IWIN2019

International Workshop on Informatics

Proceedings of
International Workshop on Informatics

September 8-11, 2019
Hamburg, Germany

Informatics Society

Sponsored by Informatics Society

IWIN2019

International Workshop on Informatics

Proceedings of
International Workshop on Informatics

September 8-11, 2019
Hamburg, Germany

Informatics Society

Sponsored by Informatics Society

Publication office:
Informatics Laboratory
3-41, Tsujimachi, Kitaku, Nagoya 462-0032, Japan
Publisher:
Tadanori Mizuno, President of Informatics Society
ISBN:
978-4-902523-46-1
Printed in Japan

Table of Contents

Session 1: Intelligent Transportation Systems

(Chair: Yoshia Saito) (9:10 - 10:25, Sep. 9)

Session 2: Multimedia and Groupware

(Chair: Katsuhiko Kaji) (10:35 - 11:50, Sep. 9)

Session 3: Network and Security

(Chair: Kozo Okano) (13:10 - 15:15, Sep. 9)

Keynote Speech 1

(15:30 - 16:10, Sep. 9)

Keynote Speech 2

(16:10 - 16:50, Sep. 9)

Session 4: Social Network and Data Analysis

(Chair: Ryozo Kiyohara) (9:00 - 10:15, Sep. 10)

Session 5: Data Models

(Chair: Takuya Yoshihiro) (10:25 - 12:5, Sep. 10)

Keynote Speech 3

(13:00 - 13:40, Sep. 10)

Session 6: Systems and Services

(Chair: Tomoki Yoshihisa) (13:50 – 15:30, Sep. 10)

Session 7: Industrial Applications

(Chair: Hiroshi Inamura) (15:40 - 16:55, Sep. 10)

Message from the General Chairs


It is our great pleasure to welcome all of you to Hamburg, Germany, for the 13th International Workshop on Informatics (IWIN 2019). This workshop has been held annually by the Informatics Society. Since 2007, the workshops were held in Naples in Italy, Wien in Austria, Hawaii in the USA, Edinburgh in Scotland, Venice in Italy, Chamonix in France, Stockholm in Sweden, Prague in Czech Republic, Amsterdam in Netherlands, Riga in Latvia, Zagreb in Croatia, and Salzburg in Austria respectively.

In IWIN 2019, 25 papers were accepted after peer reviewing by the program committee. Based on the papers, seven technical sessions were organized in a single track format, which highlighted the latest research results in the areas such as Intelligent Transport System (ITS), Network and Security, Social Network and Data Analysis, Data Models, Systems and Services, and Industrial Applications. IWIN2019 will also welcome three keynote speakers: Dr. Phil. Tim Ziemer of Researcher at the University of Bremen, Prof. Dr. Rolf Bader of Systematic Musicology at the Institute of Systematic Musicology, University of Hamburg, and Dr. Katsuhiko Kawazoe of Senior Vice President / Head of Research and Development Planning, NTT(Nippon Telegraph and Telephone Corporation). We really appreciate their participation in the workshop.

We would like to thank all the participants and contributors who made the workshop possible. It is indeed an honor to work with a large group of professionals around the world for making the workshop a great success. We are looking forward to seeing you all in the workshop. We hope you enjoy IWIN 2019.

September 2019

Masao Isshiki

Keiichi Abe

Organizing Committee

General Co-Chairs

Masao Isshiki (Kanagawa Institute of Technology, Japan)

Keiichi Abe (Kanagawa Institute of Technology, Japan)

Steering Committee

Hitoshi Aida (Tokyo University, Japan)

Toru Hasegawa (Osaka University, Japan)

Teruo Higashino (Osaka University, Japan)

Tadanori Mizuno (Aichi Institute of Technology, Japan)

Jun Munemori (Wakayama University, Japan)

Yuko Murayama (Tsuda College, Japan)

Ken-ichi Okada (Keio University, Japan)

Norio Shiratori (Chuo University / Tohoku University, Japan)

Osamu Takahashi (Future University Hakodate, Japan)

Program Chair

Yoshia Saito (Iwate Prefectural University, Japan)

Financial Chair

Tomoya Kitani (Shizuoka University, Japan)

Publicity Chair

Yoshitaka Nakamura (Future University Hakodate, Japan)

Program Committee

Chiaki Doi (NTT DOCOMO, Inc., Japan)
Fumiaki Sato (Toho University, Japan)
Gotoh Yusuke (Okayama University, Japan)
Hideyuki Takahashi (Tohoku Gakuin University, Japan)
Hiroaki Morino (Shibaura Institute of Technology, Japan)
Hironobu Abe (Tokyo Denki University, Japan)
Hirosato Tsuji
(Mitsubishi Electric Information Systems Corporation, Japan)
Hiroshi Sugimura (Kanagawa Institute of Technology, Japan)
Hiroshi Inamura (Future University Hakodate, Japan)
Hiroshi Mineno (Shizuoka University, Japan)
Hiroshi Yoshiura
(The University of Electro-Communications, Japan)
Hirozumi Yamaguchi (Osaka University, Japan)
Hisao Fukuoka (Tokyo Denki University, Japan)
Katsuhiko Kaji (Aichi Institute of Technology, Japan)
Ken Ohta (NTT DOCOMO, Inc., Japan)
Kozo Okano (Shinshu University, Japan)
Makoto Imamura (Tokai University, Japan)
Masaaki Shirase (Future University Hakodate, Japan)
Masahiko Ishino (Bunkyo University, Japan)
Masaji Katagiri (NTT DOCOMO, Inc., Japan)
Masakatsu Nishigaki (Shizuoka University, Japan)
Minoru Kobayashi (Meiji University, Japan)
Shigemi Ishida (Kyusyu University, Japan)
Shinichiro Mori (Chiba Institute of Technology, Japan)

Takuya Yoshihiro (Wakayama University, Japan)
Takuya Yuizono
(Japan Advanced Institute of Science and Technology)
Tetsushi Ohki (Shizuoka University, Japan)
Tomoki Yoshihisa (Osaka University, Japan)
Tomoo Inoue (University of Tsukuba, Japan)
Tomoyuki Yashiro (Chiba Institute of Technology, Japan)
Tsukasa Kudo
(Shizuoka Institute of Science and Technology, Japan)
Yoshiaki Terashima (Soka University, Japan)
Yoshinobu Kawabe (Aichi Institute of Technology, Japan)
Yu Enokibori (Nagoya University, Japan)
Yuichi Bannai (Kanagawa Institute of Technology, Japan)

Session 1:
Intelligent Transportation Systems

(Chair: Yoshia Saito)

A Study for Zero Emission Lifestyle at Sparse Land

Yuto Nagai* and Ryozo Kiyohara*

* Kanagawa Institute of Technology, Japan
{s1621115@cco, kiyohara@ic}.kanagawa-it.ac.jp

Abstract -There will be regulations of zero emission vehicles around the world. However, zero emission function is only applied to vehicles. In this case, much electric power is made by thermal power generation. Therefore, we have to think not only about zero emission vehicles but also about lifestyle. There are many ways for getting electric power, e.g. solar power, hydroelectric power, geothermal power, and etc. These technologies require wide space in many cases. However, if it only is focuses to vehicles, it might require not wide space. Therefore, we propose the zero emission lifestyle for mobility at Sparse Land. Then we list the require parameters and simulate it.

Keywords: zero emission, electric vehicle, ITS, autonomous vehicle, elderly person

1 INTRODUCTION

Recently, there are many problems on the earth. Especially, high temperature problem is very important issue. A main issue is CO2 problems about vehicles in the world. In the developed countries, there are many solutions for this problem (e.g. public transportation, ridesharing, electric vehicle (EV)) for zero-emission.

However, the motivation of zero-emission is very poor of only reducing the CO2. Therefore, the political mean have to be required (e.g. tax of economical vehicle, or priority lane for riding more than two people, and etc.)

The analyzing result of the behavior of drivers who use sharing the car shows that the driver who is a member of car sharing intend to use public transportation, bicycle, and etc. and they reduce the CO2 about 30 % and 157.8 [kg /year/person] as a result [1].

Adoption of EV for carsharing can reduce 80 % of CO2 [2]. The purpose of using carsharing is going to shopping mall, day trip which means going nearby areas as shown in Table 1. In this case which many people have same purpose, public transportation is effective.

Table I purpose of carsharing

g

item	Ratio (%)
Shopping to large shopping mall in suburban area	58.4
day trip leisure	56.5
Shopping at nearby shop	39.7
Driving	35.7
Transport of baggage	35.6
Picking up	34.3

Using public transportation can reduce 30% of CO2. Moreover Japanese Police recommend the elderly people to return their drivers licenses. Therefore, some local governments distribute the discount coupon for public transportations in the urban area.

However, in the local city, there are poor public

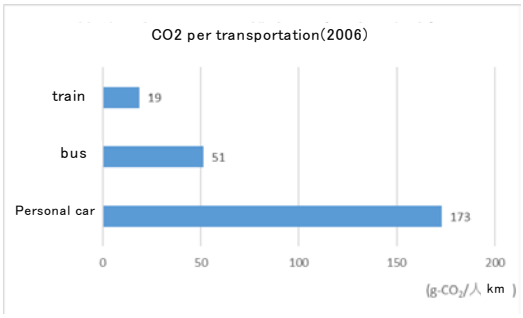


Figure 1 Co2 per transport

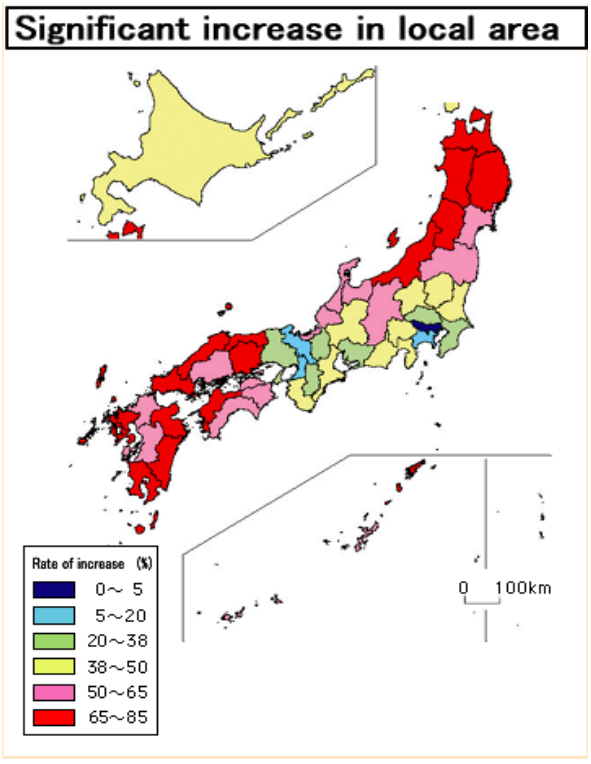


Figure 2 Transiton of Co2

transportations. Therefore, people user their own car and decrease the ratio of return the driver's license. Therefore, it cannot reduce Co2 as shown in Figure 2.

Moreover drivers for buses and taxies tends to be elder from Figure 3.

Therefore, rideshare with autonomous for safety and electric vehicle (EV) for reducing CO2 is very important. Moreover, its electric power should be supplied without CO2 as solar power.

In this paper, we study suitable number of EV and solar power equipment for typical local city. Then, we propose that how to simulate it.

2 WHAT IS RIDESHARE

We define that rideshare means a cupple of people ride a car which is personal owned car with paying its cost. The large difference from on-demand bus is that its driver is professional or not. In this paper, we define legacy rideshare as shown in Table 2.

3 RERATED STUDY

Effective method for carsharing was proposed [3]. In this study, authors evaluate their system by simulation from many aspects which are travelling distance, power charging, satisfaction of driver and etc. This evaluation assumes all vehicles for carsharing are EVs. However, in this research, they did not assume autonomous vehicles. Moreover, one of the goal is park and ride. Therefore, poor public transportation was not assumed.

Then, we should extend this study for zero emission environments with autonomous vehicles and poor public transportations.

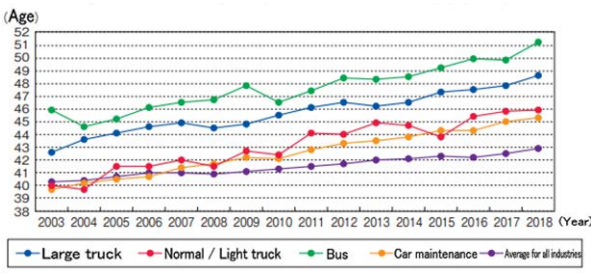


Figure 3 Average ages in workers at transportation servies

Table 2 Rideshare

	Legacy rideshare	Rideshare in this paper	Ondemand bus
Customer	Agreement with owner	Agreement with owner	request
Owner(driver)	no-professional	No-professonal	professonal
報酬	no	pay	pay

There is a study for ridesharing for the public transportation in the depopulated area [4], In this study, there are few users because there are many cases of that origin location to bus stop is very far. However, this study assumed the non-autonomous vehicles. We think that it is difficult to achieve zero-emission by only non-autonomous vehicles.

Then we assumed the ridesharing for zero-emission by autonomous vehicles.

4 PROPOSED METHOD

Many people say that EVs reduce the CO2 which is generated by vehicles. However, as the sources of electric power are shown in Figure 4, 5, about 70% of power is generated from oil. This means EVs can reduce the CO2 only 30%. This system is not estimated by many people.

Therefore, zero emission system shown in Figure 6 are ideal system.

Then, we simulate our new model in Sagamihara city which is Japanese local city (as shown in Figure 7) by multi-agent simulation. Our model includes charging time, distance of driving with no recharge,

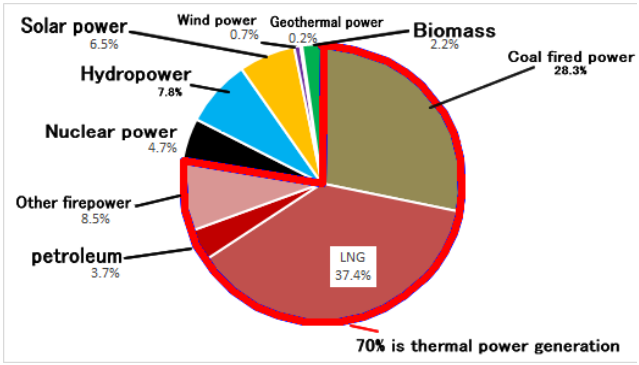


Figure 4 Ratio of origin of electric power

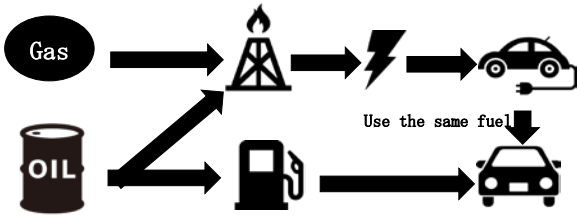


Figure 5 Legacy flow of electric power

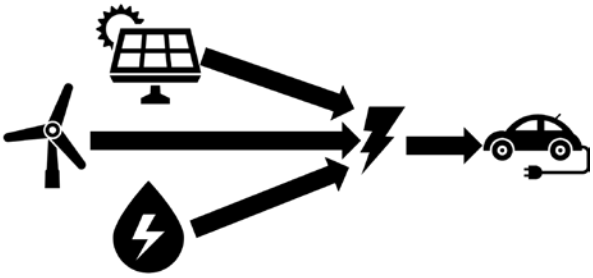


Figure 6 Proposed system

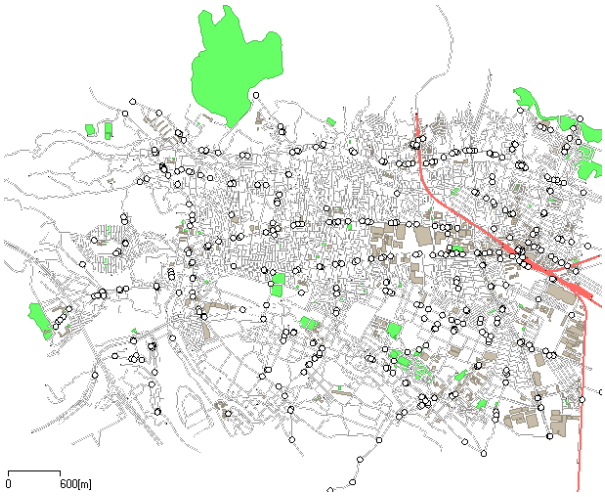


Figure 7. A part of Sagamiha City (google map)

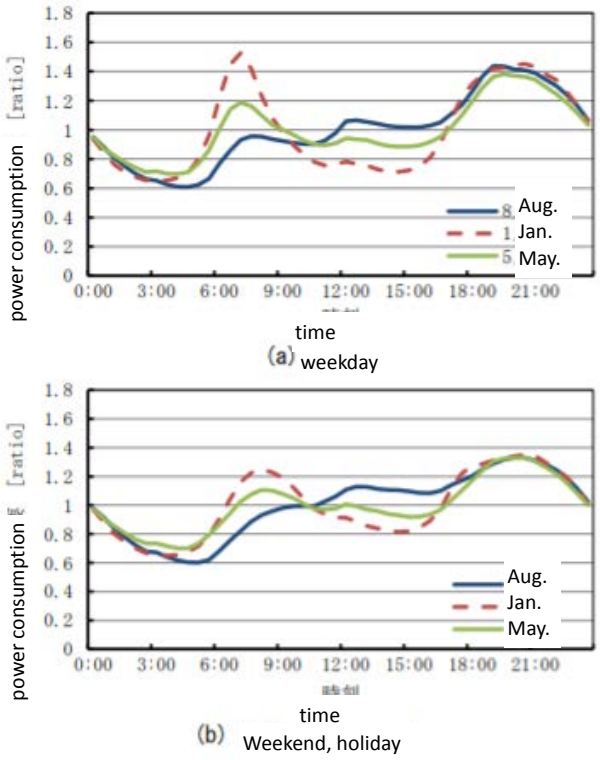


Figure 8. Average power consumptions in typical families.

Table 2. Specification of EV for simulation

manufactureare	Mitsubishi	Nissan	Nissan
Type	ZAA-HD4WLDD	ZAA-ZE0	ZAA-ZE1
Capacity of Battery	16kwh	24kwh	40kwh
Distance without recharging	164kmo	200km	400km
Normal charging	7 hours	8 hours	16 hours
Rapid charging	30 min	30 min	40 min
Price	JPY 295	JPY 377	JPY325

5 POWER CONSUMPTION IN NORMAL FAMILIES

If EV consume the energy without oil, many families might consume electric power from oil which means generating CO2.

However, there are a typical pattern of consuming electric power as shown in Figure 8. Therefore, when the electric power is not consumed so much, the power should be stored at batteries.

Then we simulate with actual EVs which are shown in Table 3. However, these data are gathered from catalogue. Therefore, we should gather the actual data from OBD port.

The simulator is multi-agent and network simulator which is Scenargie[5]. Moreover, moving model is based on [3] which are as follows.

$$F_r = \mu Mg \quad (1)$$
$$F_w = \rho CdSv^2/2 \quad (2)$$
$$F_a = Ma \quad (3)$$
$$F_g = Mg\sin\theta \quad (4)$$
$$F_d = F_r + F_w + F_a + F_g \quad (5)$$

Moreover, the vehicle model is based on [6] for elderly person. This simulation should be repeatedly executed for many parameters as shown in Table 3.

6 SUMMARY

In this paper, we have mentioned the zero-emission transporting system would be required and the environments should be assumed autonomous vehicles and EV. Moreover, the requirements are required in the local cities which are depopulation.

Table 3 many parameters in equation

Parameters	unit	contents	value
μ		Rolling resistance	Constant
M	Kg	vehicle total weight	From catalogue
g	m/s^2	Gravity acceleration	Constant
ρ	kg/m^3	Air density	Constant
Cd		Air resistance coefficient	From catalogue
S	m^2	Total projected area	From catalogue
v	m/s	Velocity	From simulator
a	m/s^2	acceleration	From actual GPS
θ	rad	slop	From map data

Then we propose the simulation should be done by many parameters based on actual data. In the near future, we gather many data and discuss about it.

REFERENCES

[1] Yamamoto T., Naruse H., Morikawa T., ”Analysis of Effects of Carsharing on Car Ownership and Travel Behavior,” The 3rd Japan-China Joint Seminar on Transportation and Urban Planning, August 4, Kumamoto, Japan, 2007.

[2] Takeuchi H., Yachi T., “CO_2 reduction by the small-size EV in car-sharing,” IEICE technical report. EE2010-2(2010-05), p37-42(2010)

[3] Shibata N., Ueda T., Sun W., Ito M., “EVTour: Online Scheduling System for Tours with Multiple Destinations by One-Way EV Sharing,” IEEE VTC FALL, 2016.

[4] Yotsutsuji H., Sasaki K.,, Yamamoto M., “Availability of Volunteer-based Dynamic Ridesharing with Bipartite Group in a Low-density Small Community in Japan,” Journal of the Eastern Asia Society for Transportation Studies, Vol. 10, pp. 1009-1024, 2013

[5] Mineo Takai, Jay Martin, Shigeru Kaneda, Taka Maeno, “Scenargie as a Network Simulator and Beyond, “ Journal of Information Processing, “ Vol.27, No.1, pp.2-9, 2019

[6] R. Malinauskas, “The Intelligent Driver Model: Analysis and Application to Adaptive Cruise Control,”. Clemson University Theses. 1934. https://tigerprints.clemson.edu/all_theses/1934, 2014

Expanding the Recognition Range for a Blind Corner

Hiroki Fujii*, Yusuke Takatori **
Kanagawa Institute of Technology, Japan
*s1612021@cco.kanagawa-it.ac.jp, **takatori@ele.kanagawa-it.ac.jp

Abstract—In this paper, an ITS application that detects obstacles in a blind area in a road curve section is proposed, and the recognition range expansion performance of the application is analyzed. It is assumed that the proposed application obtains information of obstacles in a blind area with an image reflected on the side panel of the preceding vehicle running on the same lane by a stereo vision camera. Furthermore, this paper presents a method to analyze the range expansion in which the invisible opposite lane following the back of the curve section can be recognized. From the analysis results, it is suggested that there is a possibility of expanding the recognition range to the invisible area following the back of the curve section.

Keywords: ITS, Stereo vision camera, obstacle detection

1 INTRODUCTION

Intelligent transport system (ITS) is a system that advances the movement of people and objects using information communication technology. Research on driving assistance systems and automated driving is have been studied to improve road traffic safety, with these technologies, the vehicle acquires information on surrounding obstacles to realize recognition, judgment, and operation support that enhances safety. Nowadays, acquisition of obstacle information within the direct view of vehicles has been realized by in-vehicle obstacle sensors (Laser range finder / Millimeter waves / Vision systems etc.).

On the other hand, detecting obstacles using vehicular on-board sensors that cannot be seen from a vehicle is difficult. As means for acquiring obstacle information in the blind area, systems using road-vehicle communication and inter-vehicle communication have been studied [1][2]. However, road-to-vehicle communication can only be used at the coverage area of the roadside unit. In addition, since inter-vehicle communication requires that both the ego-vehicle and the other vehicles are equipped with on-vehicle devices, the system performance depends on the penetration rate of on-vehicle devices. Therefore, there is a need for a technology that can obtain obstacle information in a blind area by the on-board sensors without depending on the location.

There are not many reports on technology for acquiring out-of-sight obstacle information using an on-vehicle sensor. Patents have been reported that use a millimeter wave radar to detect an obstacle ahead of a leading vehicle via the bottom of the leading vehicle and between the road surface [3]. In this technique, since a sensor for forward obstacle detection is used, the area which can be sensed is limited. In addition, the millimeter wave radar is limited to the information in the depth direction of the obstacle and has a low angle resolution. Many of the existing automatic braking systems use a vision sensor to recognize pedestrians, road conditions, and signs. It is also desirable to acquire information other than position information even in the case of obstacle detection in a blind area.

The authors have previously studied a method for estimating the position of an obstacle in a blind area that assumes the use of a stereo camera [4][5]. In these studies, methods of estimating the position of an obstacle from a virtual image of the obstacle reflected on a vehicle traveling in the adjacent lane are considered. Currently, the authors aim to increase the number of targets for obtaining reflection images in order to make wider use of the basic concepts examined in the previous research.

In this paper, the authors consider acquiring and utilizing blind area information reflected in the side panel of a leading vehicle that travels a blind curve and propose an application system and analyze the performance of it.

This paper is organized as follows. Section 2 explains the position estimation of an obstacle in a blind area by a stereo vision system. In Section 3, An application for providing driving assistance information by blind area obstacle detection in a curve is proposed. In Section 4, the method of performance analysis is provided, and the analysis result is shown. Finally, in Section 5, we conclude this paper.

2 POSITION ESTIMATION IN BLIND AREA WITH STEREO VISION CAMERA

2.1 Principle of the position estimation of an obstacle in blind area [4]

Stereo vision camera is a camera system that two cameras with the same specifications are fixed at a predetermined distance apart and directed in the same direction. In the images taken by each camera, the position of an obstacle is shifted between the left and right images; the difference is called “disparity”. The stereo vision camera calculates the depth (distance between the camera and the obstacles) of the obstacles by the disparity. Such systems are already widely adopted in commercial vehicles.

By facing the stereo camera to a flat mirror, the distance to the virtual image in the mirror world can be obtained. Figure 1 shows the principle of the position estimation of an obstacle in blind area. Then, if the state of the mirror surface to the stereo vision camera (relative position and angle, etc.) is known, it is possible to estimate the virtual position of the obstacle in the mirror world. Moreover, the position of the real image is estimated by folding the virtual position against the mirror surface. Consequently, if a real image cannot be seen directly from the stereo vision camera, it is possible to estimate the position of an obstacle in blind area if the reflection image can be seen on the mirror surface. In particular, when a large vehicle exists around the ego-vehicle, there are more cases where the direct viewable area is limited,

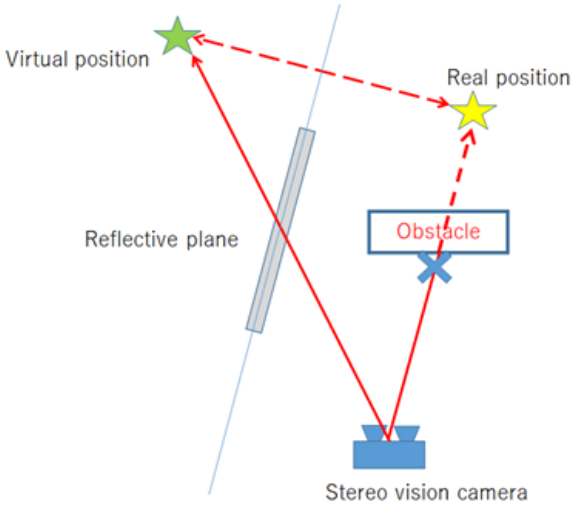


Figure 1. Principle of the position estimation of an obstacle in blind area.

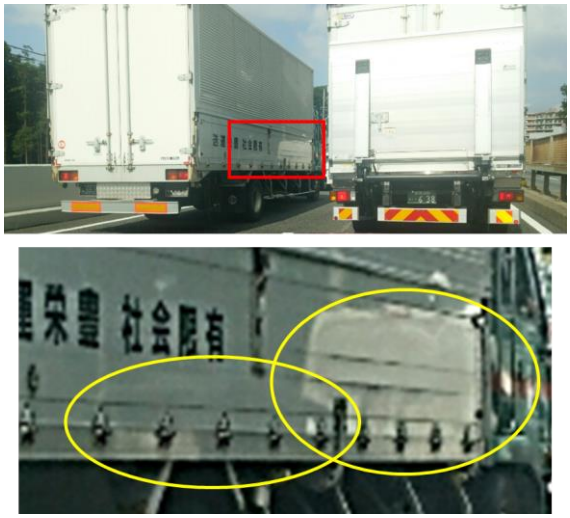


Figure 2. Reflected image of non-line-of-sight obstacles on the side panel of the vehicle running in the adjacent lane

so it is effective to utilize such an environment to obtain an obstacle information in a blind area.

2.2 Example of Application [4][5]

Previous studies [4][5] assume a road environment with multiple lanes in the same direction. These studies present an application that detects an obstacle placed in front of the preceding vehicle of the ego-vehicle from images reflected on the side panel of the vehicle running in the adjacent lane. Figure 2 shows the image that a non-line-of-sight vehicle placed on the placed in front of the preceding vehicle of the ego-vehicle reflected on the side panel of the vehicle running in the adjacent lane. In the ref. [4], the basic principle of the position estimation of the obstacle that cannot be seen directly is established on the assumption of the side panel of a large vehicle with a box-shaped carrier, and in the ref. [5], the position estimation method that uses a semitransparent reflective surface such as a grass of a vehicle window as a reflective surface has been proposed. While this application

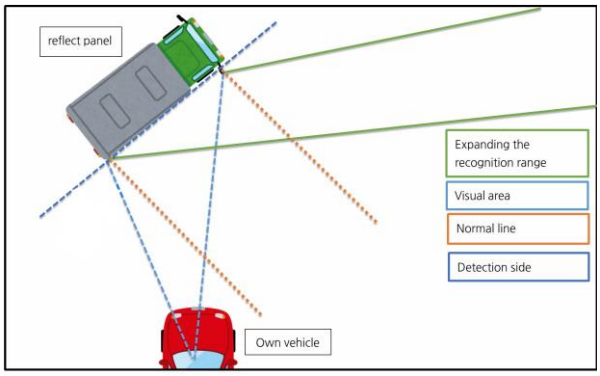


Figure 3. Obstacle positioning application for a blind corner

deals with the reflective surfaces of vehicles in adjacent lanes, methods and applications that apply this principle to other reflective phenomena that can be acquired while the vehicle is traveling are not considered.

3 OBSTACLE POSITIONING APPLICATION FOR A BLIND CORNER

On straight roads, the preceding vehicle is not valid target for above-mentioned obstacle position estimation method. This is because most of information obtained from the scenery reflected on the rear panel of the preceding vehicle is the information within the line of sight of the ego-vehicle. Therefore, there is little chance of getting information of blind obstacles. Meanwhile, in the case of a curved road, when the heading of the preceding vehicle differs from that of ego-vehicle, ego-vehicle may be able to see the side panel of the preceding vehicle. Then obstacles in the blind area may be reflected on the side panel of the preceding vehicle. Therefore, as shown in the Figure 3, the authors propose an application that uses the reflection surface of the preceding vehicle to estimate the position of obstacles present in the oncoming lane in a blind area that cannot be seen from the ego-vehicle. The overview of the application is as follows:

- Ego-vehicle is equipped with a stereo vision camera in the center of the front.
- Ego-vehicle recognize the status of the side panel of the preceding vehicle by the stereo vision camera or other obstacle detection sensors.
- The application system obtains two images from the stereo vision camera and process a matching of both images to estimate the distance and direction from the ego-vehicle to the virtual image.
- The virtual image position is mapped on the ego-vehicle center coordinate system by the distance and direction from the ego-vehicle to the virtual image.
- The position of the obstacle is estimated by folding back the position of the virtual image with respect to the reflecting surface in consideration of the state of the surface of the side panel.

The expanded recognition range depends on the positions of the camera of ego-vehicle and the positions of the front end

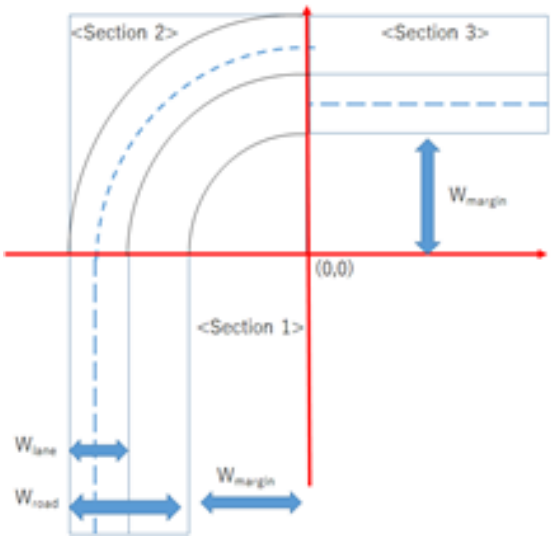


Figure 4. Road model

and the rear end of the side panel of the preceding vehicle. Therefore, it is important to assess these areas and evaluate their feasibility prior to the application development. The next section describes methods for analyzing performance and shows the result of analysis.

4 PERFORMANCE ANALYSIS

4.1 Environment

(1) Road model
Figure 3 shows the road model of this analysis. The road is a L-shaped corner and it has one lane on each side. The road model is divided into three section: Straight entrance section (Section 1), Curved section (Section 2) and Straight exit section (Section 3). The length of each straight section (Section 1 and 3) is variable and it is set to an appropriate value for analysis. The curved section (Section 2) is designed as a quarter circle. The distance from the center of the circle to the inner edge of the road is a parameter and is given by the letter W_{margin} in Figure 4. The part of inside of the quadrant with radius W is the off-road area that blocks the direct view from vehicles. This road model is assumed as a local road with less traffic, and its parameters are set according to the law on the road. The lane width (indicated by the letter W_{lane}) is set at 3 m on one side and 6 m in the width of two lanes (indicated by W_{margin}).
(2) vehicle model
The sizes of the ego-vehicle and the preceding vehicle in this analysis are described. The size of the ego-vehicle is not specified. However, it is assumed that a stereo vision camera is mounted at the center of the nose of the ego-vehicle. The preceding vehicle is assumed to be a rectangular as a large vehicle. The size of the preceding vehicle is set based on the size of a commercially available bus or truck. The size of the preceding vehicle is set as 10 meters in length and 2 meters in width, the length of the side panel of the preceding vehicle is equal to the length of the preceding vehicle. The side panel assumes a flat, undistorted panel. That is, it is possible to

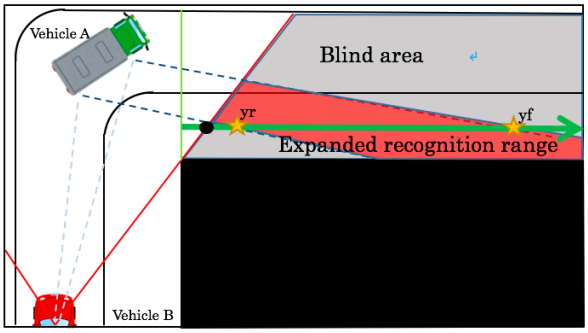


Figure 5. Application operating environment

acquire an image of an obstacle without distortion. This model is considered in a two-dimensional plane.

The initial position of each vehicle is described. The ego-vehicle is placed in section 1 and its position is set at an arbitrary value. This analysis assumes about 400 vehicle per hour as the traffic volume. And from this traffic volume, the average headway-time-distance(s) is calculated. This is multiplied by the traveling speed to obtain the average headway distance(m). The initial position is determined using this average headway-distance as the location of the ego-vehicle, the preceding vehicle is placed its center on the boundary between section 1 and section 2. It is assumed that these vehicles move at the center of the lane at a constant speed. In this analysis, a curve section limited to low speed (30 km / h) is assumed. And, the traveling speed is assumed to be 5 m / s (18 km / h).

(3) sensor model
This application requires stereo vision camera on the ego-vehicle. Moreover, about the sensor for estimating the state of a preceding vehicle side panel, as long as the necessary information can be obtained, any type of sensor mounted can be used. In the analysis, installation of an ideal obstacle detection sensor is assumed, that is, there is no limitation on the positioning distance and the field of view, and it is assumed that information within the line of sight can be obtained. The stereo vision camera is assumed to be capable of estimating the position of an obstacle that appearing on the reflective surface.

4.2 Analytical method

In order to analyze the expansion of the recognition range in a blind area by a stereo vision camera mounted on the ego-vehicle, an application operation environment shown in Figure 5 is constructed. The analytical method is as follows. First, a reflecting surface which can be formed when the preceding vehicle enters at a corner and turns is determined. Next, using the camera position of the ego-vehicle and the positions of the front end and the rear end of the side panel of the preceding vehicle, the area to be reflected on the panel is determined. The area reflected on the panel is the area between the two reflected lights; One is the reflected light for the light ray incident from the camera to the front end of the side panel and another is the reflected light for the light ray incident to the rear end of the side panel. Therefore, these two reflected light rays are treated as straight line equations on the

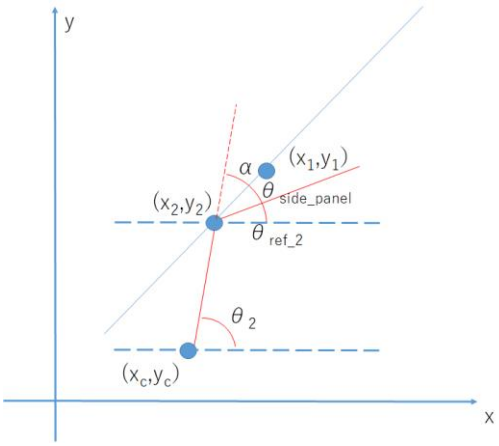


Figure 6. Position coordinate of the preceding vehicle in Section 2.

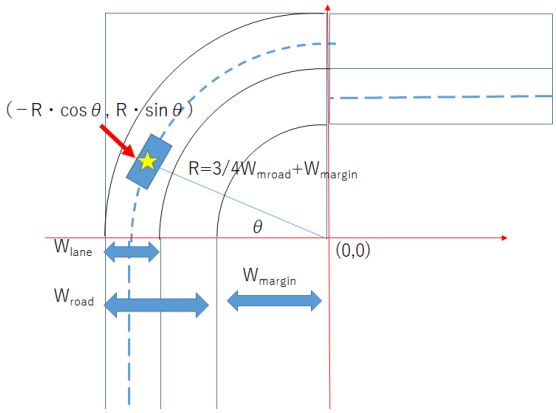


Figure 7. Position coordinate of the preceding vehicle in Section 2.

coordinal plane, and the area between the two straight lines is the area where the overlapping area with the opposite lane area of the blind area is the expansion area of the recognition range.

Here, the equation of the reflected light with respect to a ray incident from the camera to the rear end of the side panel is derived. The coordinate is set as shown in Figure 6. The coordinates of the front and rear end of the panel are set (x_1, y_1), and (x_2, y_2) , respectively. The coordinates of the the sterel vision camera is set (x_{2c}, y_c). The equation from the camera to the rear side is:

$$y = \frac{y_2-y_c}{x_2-x_c}x + \frac{x_2y_c-x_cy_c-y_2+y_c}{x_2-x_c} \quad (1)$$

Therefore, the angle θ_2 between the x-axis and the straight line of equation (1) is given by the following equation from the slope of the straight line of equation (1)

$$\theta_2 = tan^{-1}\left(\frac{y_2-y_c}{x_2-x_c}\right) \quad (2)$$

Subsequently, the equation of the straight line of the side panel is derived from the front and rear coordinates.

$$y = \frac{y_2-y_1}{x_2-x_1}x + \frac{x_2y_1-x_1y_1-y_2+y_1}{x_2-x_1} \quad (3)$$

Therefore, the angle θ_{side_panel} between the x axis and the straight line of Equation (3) is given from the slope of Equation (3).

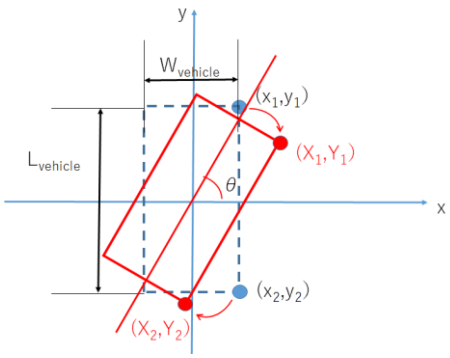


Figure 8. Position coordinates of the front end and rear end on the road coordinate system

$$\theta_{side\_panel} = tan^{-1}\left(\frac{y_2-y_1}{x_2-x_1}\right) \quad (4)$$

Using these angles, the angle α between equation (1) and equation (3) is given by the following equation.

$$\alpha = \theta_2 - \theta_{side\_panel} \quad (5)$$

From these, the inclination of the equation of the ray reflected at the side rear end is given by :

$$a_{ref\_2} = \tan(\theta_2 - 2\alpha) \quad (6)$$

Since this passes through the side rear end, the ray equation is given by:

$$y = a_{ref\_2}x + y_2 - a_{ref\_2}x_2 \quad (7)$$

Similarly, the equation of the reflected light of the ray incident from the camera to the front end of the side panel is also determined.

Furthermore, in order to obtain the coordinates of two points of the preceding vehicle traveling on the curve section, the following calculation is performed. The preceding vehicle enters a curve at time t = 0. The moving distance to the time t after entering the curve section is calculated by v×t. Since this corresponds to the length of the arc, the movement angleθ after entry is θ = vt / R. Therefore, the center coordinate of the preceding vehicle (x_{center},y_{center}) is derived as follows.

$$(x_{center},y_{center}) = (-R\cos\theta,\ R\sin\theta) \quad (8)$$

These relation is shown in Figure 7. In addition, since the vehicle moving along the arc rotates -θ with respect to the y-axis, the position coordinates of the side front and rear ends viewed from the vehicle coordinate system rotates -θ with respect to the origin(shown in Figure 8). The position coordinates of the side front and rear end viewed from the vehicle coordinate system are given by

$$(X_k, Y_k) = (x_k cos\theta + y_k sin\theta, -x_k sin\theta + y_k cos\theta) \quad (9)$$

By adding equation (9) to equation (8), position coordinates of the front and back sides of the side on the road coordinate system can be obtained. Moreover, the positions of x at which the obtained two straight lines intersect with the straight line of y = margin + 1/4 width are determined as x_f and x_r, respectively.

The straight line separating the area which can be seen from the camera and the area which cannot be seen is a straight line which touches a circle of radius margin [m].
The position of x that intersects this straight line and the straight line of y = margin + 1 quarter lane width is determined as x_{blind} and compared with x_f and x_r.

4.3 Result

The extension area of the recognition range when the ego-vehicle approaches the corner following the preceding vehicle was analyzed. Figure 9 shows the values of x_f , x_r and x_{blind} with respect to elapsed time t (s), respectively. x_f or x_r are the intersection of the reflected light at the front or rear end of the side panel and the center line of the opposite lane of section 3. x_{blind} is the intersection of the reflected light at the front or rear end of the side panel and the center line of the opposite lane of section 3. Moreover, x_{blind} is the intersection of the tangent of the ego-vehicle's camera to the arc inside the road and the center line of the opposite lane in section 3. From this result, x_r becomes positive between t=3 and t=4 seconds. In this period, the following vehicle can recognize the opposite lane of the blind corner for the first time. Moreover, it is shown that the values of x_f and x_r increase with the elapsed time. On the other hand, x_{blind} is smaller than either x_f or x_r. That is, the recognition range penetrates from the place near the corner to the back of the corner as the time elapsed.

Meanwhile, x_f and x_r are drastically changed around t = 6 seconds. In this period, x_f and x_r are temporarily negative, and the blind area cannot be recognized. This is because the positional relationship between x_f and x_r is switched depending on the relationship between the panel front end and rear end positions of the preceding vehicle and the coordinates of the lane center line of the opposite lane of the blind area.

When the y-coordinate of the preceding vehicle is smaller than the coordinates of the lane center line of the opposite lane, the recognition range is determined as shown in Figure 10. On the other hand, when the y-coordinate of the preceding vehicle is larger than the coordinates of the lane center line of the opposite lane, the y-coordinate of the preceding vehicle is smaller than the coordinates of the lane center line of the opposite lane (See in Figure 11). The slopes of the respective straight lines become minute or negative just at these switching times (about 4 to 7 seconds), and there are cases where both x_f and x_r become negative. However, after x_f and x_r interchange occurs, it can be seen that the recognition area of the blind area is expanded as before the interchange.

It has been shown that although it cannot be recognized temporarily between 4 seconds and 7 seconds, it is possible to recognize an obstacle in a blind area about 30 m from a corner before 4 seconds. Besides, the distance traveled by the vehicle present in the blind area in three seconds is estimated to be about 15 m. This means that there is a margin of 15m to the corner when the cognitive function is restored even if a vehicle farther than 30m in the opposite lane approaches a period when the blind area cannot be recognized. Consequently, during periods when the blind area cannot be recognized, it is desirable to provide driving assistance using information already obtained.

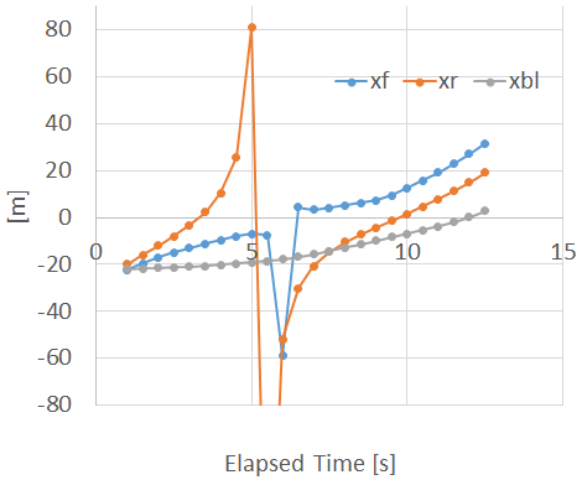


Figure 9. extension of the recognition range

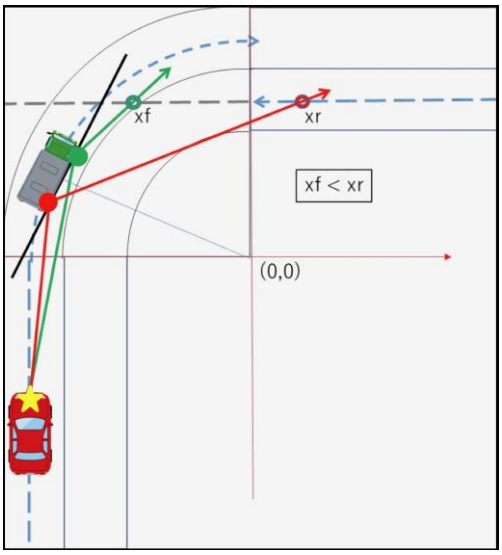


Figure 10. Position coordinates of the front end and rear end on the road coordinate system

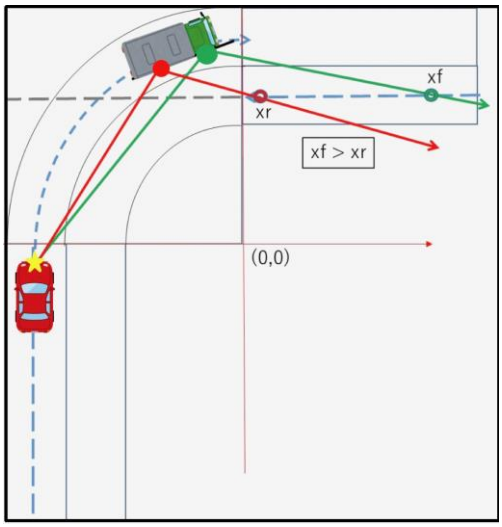


Figure 11. Position coordinates of the front end and rear end on the road coordinate system

5 CONCLUSION

In this paper, an application that detects obstacles in a blind area from a reflected image on the side panel of a preceding vehicle that turns in a curve is proposed. Furthermore, we

described analytical method of the application performance and analyze it on a L-shaped corner situation. The analysis result indicated that as the vehicle approached the corner, the recognition range was extended from the front of the blind area to the back. In addition, in the environment assumed this study, while approaching the corner, recognition of the blind area became temporarily impossible. However, it is a short-term decline in cognitive function, and there is a possibility that support can be provided until the functional decline by using information obtained so far.

In this analysis, some set parameter values are given as fixed values. It is necessary to change the values of these parameters to perform a wider analysis. In addition to theoretical analysis, it is necessary to study using an actual system. Future subjects are development of stereo vision camera system corresponding to real environment and performance evaluation in real environment.

REFERENCES

[1] M. Tamura, S. Takahashi, S. Yasuhara, M. Kojima, K. Minegishi, "Development of intersection safety support systems using vehicle-to-vehicle communication", 12th ITS World Congress 2005, CD-ROM (7 pages)

[2] Yusuke Takatori, Hideya Takeo, "Analysis of Vehicle Information Sharing Performance of an Intersection Collision Warning System", IEICE TRANS. FUNDAMENTALS, VOL.E100–A, NO.2 FEBRUARY 2017.

[3] Noriko Shimomura, JP Patent 2005-182137, 2005.7.7.

[4] Riku Kawamoto, Yusuke Takatori," A study on position estimation of obstacles reflected in mirror",IEICE TR, vol. 118, no. 79, ITS2018-1, pp. 1-6, 2018.

[5] Sakuma Nakamura, Yusuke Takatori," A Study on Estimation of Obstacle Position Using Reflected Image on Transparent Surface", IEICE TR, vol. 118, no. 79, vol. 118, no. 464, ITS2018-92, pp. 45-50, 2019.

A Method for Estimating Road Situations Based on Bicycle Behavior Sensing towards Recommendation of Comfortable Route

Yoh Shiraishi [*], and Yusuke Takeyasu[**]

[*]School of Systems Information Science, Future University Hakodate, Japan
[**]Graduate School of Systems Information Science, Future University Hakodate, Japan
siraisi@fun.ac.jp

Abstract - In recent years, the number of bicycle users has increased due to health promotion and consideration of environmental issues. On the other hand, road maintenance for bicycle users is not sufficiently advanced. There is a problem that bicycle users cannot ride comfortably. If the type and scale of obstacles (drainage covers, parked vehicles, etc.) on roads and road surface condition such as cracks and holes can be estimated, recommendation of comfortable routes for bicycle users will be realized. This study proposes a method for estimating road situations (obstacles and road surface conditions) by sensing bicycle behaviors. We assume the estimated results are shared among bicycle users by participatory sensing approach. In the proposed method, a smartphone is mounted on a bicycle, and sensor data such as acceleration and angular velocity from the smartphone are collected as bicycle behavior data. The proposed method extracts feature values from the characteristic behavior during obstacle avoidance, and classifies the avoided obstacles based on the extracted features. Also, it estimates road surface condition based on the acceleration data in the vertical direction. In the evaluation experiment on obstacle estimation, the estimation accuracy improved and the overall F-measure was 0.88 under the classification criteria based on the scale of avoidance behavior.

Keywords: behavior sensing, obstacle avoidance behavior, road situation estimation, route recommendation for bicycle users, participatory sensing

1 INTRODUCTION

In recent years, the number of bicycle users has increased in Japan due to health promotion and environmental issues. Along with the increase of bicycle users, lots of the accidents between bicycle users and pedestrians frequently occurred. The road traffic law has been revised in order to reduce these accidents [1]. Under the revised law, bicycle users are required to move on the left side of a road in principle. However, it is hard to say that road maintenance for bicycle users is sufficiently advanced. There are many roads that prevent bicycle users from comfortable riding on a bicycle: a road with a narrow roadside belt that cannot secure a safe distance from a vehicle traveling sideways when avoiding obstacles such as drainage lids, a road where roadside belts and bicycle lanes are blocked by parked vehicles, and a road with unevenness due to road surface damage. Consequently, there is a problem that bicycle users cannot ride comfortably when traveling on roads according to the revised traffic rules.

One approach to solving this problem is to collect information about which roads are suitable for traveling by a bicycle and provide these information to bicycle users. In an existing service [2] based on this approach, each bicycle user collects information about the comfortability (“easy to ride”, “hard to ride”, etc.) of the routes where he or she traveled, and shares it with other users. However, in this service, users need to input the information manually. For this reason, it takes much time for a user to input the information, and it is difficult for a user to input the information on all the roads that the user traveled by its bicycle. There is a drawback that it is not possible to collect information about roads comprehensively for bicycle users.

In order to solve this drawback, this study proposes a method to automatically collect and share information on all roads on which bicycle users traveled by its own bicycle. In the proposed method, a smartphone are mounted on each bicycle, and the behavior data of the traveling bicycle are collected from lots of bicycle users based on participatory sensing approach. Participatory sensing [3], [4] is a framework in which lots of collaborators (participants) dispersed in the targeted area for sensing share data measured by their owned devices such as sensors and smartphones. Information collection based on participatory sensing is suitable for collecting information on roads for bicycle users because it can collect information over a wide range comprehensively.

This study aims to recommend a route on which a bicycle user can travel comfortably. To achieve this goal, we examine a method to estimate road situations such as the existence of obstacles on roads and the unevenness of roads that effect comfortability of the recommended route.

First, we focus on obstacles on roads, such as parked vehicles and drainage lids. We estimate information on an obstacle by using sensor data of bicycle behaviors when avoiding obstacles. Since the obstacle avoidance behaviors are considered to be different depending on the type and scale of the avoided obstacles, we can collect information about the obstacle by detecting the avoidance behaviors. When a bicycle user avoids an obstacle, it is considered that there is a characteristic tendency in the operation of turning the steering wheel and the operation of acceleration/ deceleration. Our method extracts the features of changes in steering wheel angle and the features of changes in velocity from bicycle’s behavior data, detects the obstacle avoidance behavior by machine learning by using the extracted features, and classifies the type and scale of obstacles.

Next, we focus on to road surface conditions such as unevenness of roads due to road damages and cracks. Our method estimates road surface conditions based on the characteristics of the vertical vibration of a bicycle that changes depending on the road surface. When the road surface condition is good, the vibration in the vertical direction of the bicycle will be small, but when the condition is bad, the vibration will be large. Our method also estimates the comfortability of roads based on the estimation results of global and local surface conditions.

This paper is organized as follows. We mention the related works in Section 2, and describes the details of the proposed method in Section 3. After we discuss the effectiveness of the proposed method through the experimental results in Section 4, we conclude in Section 5.

2 RELATED WORKS

In this section, we mention researches on sensing for bicycle behaviors and researches on behavior detection based on bicycle sensing data as the related works.

2.1 Sensing for Bicycle Behaviors

There are several researches [5], [6], [7], [8] on detection of bicycle behavior using multiple sensors with different type, and researches [9], [10], [11] on the behavior detection using smartphones.

First, the researches [5], [6] that use multiple dedicated sensors with different type collect information on the speed of a bicycle by using reed switches and photoelectric sensors installed on the bicycle. In addition, it collects information on the steering wheel angle of the bicycle by using a rotary encoder. In the reference [7], it collects information on the number of rotations of the pedal by using a magnetic sensor. In the reference [8], it collects information on the steering angle and the vehicle body tilt by using an angular velocity sensor because bicycle users turn not only by steering but also by tilting the vehicle body. As described above, it is possible to collect various kinds of information on the behavior of the bicycle with high accuracy by using dedicated sensors. However, in participatory sensing, it is necessary to reduce the introduction cost in order to make more bicycle users participate. Therefore, these methods using multiple dedicated sensors that are rarely possessed by general bicycle users is not considered suitable for our study.

Next, the researches [9], [10], [11], using a smartphone with several kinds of sensors collect information on the steering angle of the handle and the inclination of the vehicle body by using an angular velocity sensor in smartphone mounted on the handle. In the reference [10], it uses an acceleration sensor on the smartphone, and collects information on acceleration/deceleration/stop/running of the bicycle. In the reference [11], it collects the steering angle and the inclination of the vehicle body by using a geomagnetic sensor in the smartphone. Since smartphones are equipped with a wide variety of sensors in addition to the sensors used in the above research, a single smartphone can collect a many kinds of information. Since smartphones have become popular in recent years, and many bicycle users possess them, the introduction cost can be reduced. Therefore, a smartphone can be considered as a sensing device suitable for this study based on participatory sensing approach.

2.2 Behavior Detection based on Bicycle Sensing Data

As researches on behavior detection using bicycle's behavior data, there are several researches using threshold judgment [8] and researches using machine learning [6], [9], [10].

First, the study [8] using threshold judgment sets a threshold in the angular velocity value, and detects the behavior when avoiding the damaged part on the road surface. However, when we collect avoidance behavior data from multiple bicycle users, there can occur individual differences in the timing of starting avoidance, how to turn the handle, and how to tilt the vehicle body. Therefore, it is considered difficult to set a uniform threshold in the steering angle of the steering wheel and the inclination of the vehicle body. Also, because bicycles have a lot of wobbling, the threshold for detecting the avoidance behavior may be exceeded.

Next, the researches using machine learning [6], [10] extract feature values from bicycle's behavior data collected by sensors, and classify the bicycle behaviors when meandering riding or avoiding a street parking vehicle by decision tree as a machine learning method with the extracted features. Also the reference [9] classifies behaviors such as right/left turning and running on left/right curve by SVM (Support Vector Machine). By collecting behavior data from a plurality of bicycle users and performing machine learning using features extracted from these data, it is considered possible to flexibly cope with individual differences among bicycle users. Since a method for behavior classification using machine learning can take into account individual differences in the timing of starting avoidance or in how to turn a handle, it is considered suitable for this study based on participatory sensing.

3 PROPOSED METHOD

3.1 Research Tasks and Approaches

This study aims to estimate road situations of based on bicycle behavior sensing by using a smartphone, and especially focus on estimation of obstacles on roads and road surface conditions. We describe the tasks of this study to achieve this goal and the approaches as follows.

Task1: Collecting data with low introduction cost

It is necessary to comprehensively grasp the situations of a wide range of routes in order to recommend comfortable routes to bicycle users. This study introduces participatory sensing, and collects bicycle behavior data related on wide range of roads from lots of bicycle users. In order for more bicycle users to participate in the sensing activities, it is necessary to reduce the introduction cost so that anyone can participate. This study uses a smartphone as a sensor device

in order to reduce the introduction cost. Also, since a smartphone is equipped with a wide variety of sensors, we can collect bicycle behavior data effective for estimating road situation by installing a smartphone on each bicycle.

Task2: Extracting features useful for detecting obstacle avoidance behavior

This study focuses on the characteristic avoidance behavior when a bicycle avoids obstacles in order to detect the type and scale of the obstacles on roads by using bicycle behavior data. We focus on the steering angle and the velocity of the bicycle of bicycle behavior data in order to extract feature values effective for detecting obstacle avoidance behaviors. When a bicycle goes on straight, these two values may be relatively stable. On the other hand, when avoiding obstacles, it is considered that large change in steering wheel operations and decelerations can be observed. Therefore, we detect the obstacle avoidance behaviors by extracting the characteristic changes from the steering angle and velocity of the bicycle during obstacle avoidance.

Task3: Estimating road conditions based on bicycle behavior data

We focus on the vertical vibrations of a bicycle traveling on the road in order to estimate the road surface condition. We consider that road surface conditions can be classified by extracting features from the vertical vibration of the bicycle that changes depending on the road surface condition. For example, the vibration in the vertical direction is reduced on a smooth road surface, but is increased on a rough road surface. Also, the road surface condition can be classified into a global condition indicating the road surface materials such as asphalt and a local condition indicating cracks and holes.

3.2 An Overview of the Proposed System

This study focuses on estimation of road situation towards recommendation of comfortable routes for bicycle users. In this paper, we propose a method for obstacle detection and for road condition estimation that are considered to be highly important. Our method consists of a learning phase and an estimation phase because we use machine learning for obstacle detection and road condition estimation. Figure 1 shows an overview of the proposed system.

In the learning phase, bicycle behavior data collected from multiple bicycle users are input to the system, and multiple features used for road situation estimation are extracted from the behavior data. The extracted feature values and the corresponding correct answer labels are stored in the learning DB. In the estimation phase, it extracts features from the bicycle behavior data collected by a bicycle user, and estimates the situations of the traveling road by using the learning model constructed in the learning phase.

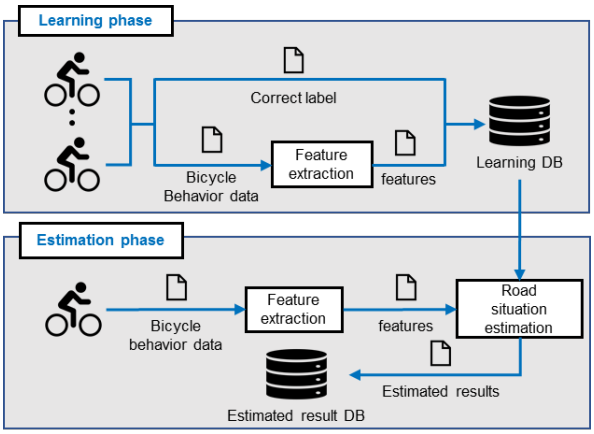


Figure 1: An overview of the proposed system

3.3 Collecting Bicycle Behavior Data

The proposed method collects bicycle behavior data by participatory sensing with multiple bicycle users. We use a smartphone as a sensing device. The smartphone was mounted on the bicycle handle as shown in Figure 2. In this figure, the smartphone is installed parallel to the ground, and the top of the smartphone is facing the traveling direction of the bicycle. The X, Y, and Z axes of the smartphone correspond to the left-right direction, front-rear direction, and vertical direction of the bicycle, respectively. We collected acceleration, angular velocity, geomagnetic data, and GPS as time series sensor data. The sampling rate is set to 100 Hz for acceleration, angular velocity, geomagnetic data, and 5 Hz for GPS, respectively. In addition, this study uses a speed sensor dedicated for a bicycle to obtain the bicycle velocity with high accuracy. The speed sensor was installed on the rear wheel of the bicycle as shown in Figure 3.

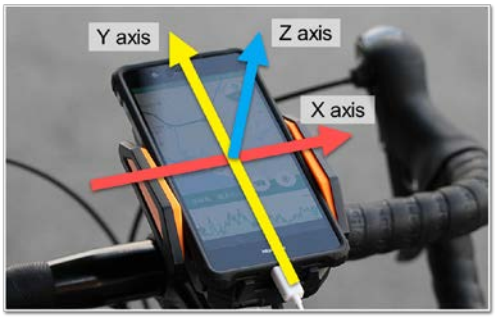


Figure 2: The setting of a smartphone

Figure 3: The setting of a speed sensor

3.4 Feature Extraction for Obstacle Detection

We focus on sensor data in the steering angle and the azimuth which changes characteristically when a bicycle avoids obstacles, and extract the feature values for classifying the obstacles from these sensor data. The feature values extracted from the azimuth and velocity data are described below.

Feature values extracted from azimuth data

As feature values for detecting obstacle avoidance behavior, we focus on large change in the azimuth angle during obstacle avoidance, and the time required for the avoidance. When a bicycle goes on straight, the traveling direction of the bicycle does not change before and after the obstacle avoidance. Consequently, it is considered possible to detect the avoidance behavior by detecting the points where the change of the azimuth angle to the positive and negative becomes maximum and minimum, respectively. However, the azimuth is always decreasing (increasing) in the case that the road is curved left (right), it cannot detect the avoidance behavior only by detecting the above two points. We introduce a linear approximation curve that approximates the tendency of fluctuations in the azimuth so that it can be applied to both when going on straight and traveling on a left/right curve road. By using the difference between the measured value and the approximate curve, it becomes possible to detect the change in the value during avoidance behavior when traveling on the left and right curves in the same way as when going on straight. The feature values in the azimuth for detecting the obstacle avoidance behavior are shown in the following equations (1) to (3).

$$Diff_{az\_max} = y_{az\_max} - f_{az}(T_{az\_max}) \quad (1)$$

$$Diff_{az\_min} = y_{az\_min} - f_{az}(T_{az\_min}) \quad (2)$$

$$T_{az\_int} = \left| T_{az\_min} - T_{az\_max} \right| \quad (3)$$

Let T_{az_max} and T_{az_min} be the time when the difference between the measured value in the azimuth and the value approximated by the linear approximate curve f_{az} (x) is the maximum (positive value) and the minimum (negative value) respectively. $Diff_{az\_max}$ represents the difference between the measured value y_{az_max} at T_{az_max} and the approximate value f_{az} (T_{az_max}). $Diff_{az\_min}$ represents the difference between the measured value y_{az_min} at T_{az_min} and the approximate value f_{az} (T_{az_min}). T_{az_int} represents the elapsed time from T_{az_max} to T_{az_min}, and is expected to be long as the avoided obstacle becomes large. Figure 4 shows these feature values with the measured data and the approximate curve.

Feature values extracted from velocity data

The results in the preliminary experiment indicate the possibility that a bicycle may rapidly slow down when avoiding an obstacle. Therefore, we focus on this tendency, and also use feature values related to the bicycle velocity data. First, as in the case extracting the feature values related to the azimuth angle, it calculates a linear approximation curve of the velocity data. Next, it calculates the maximum and minimum value of the difference between the measured value and the approximate curve value, and uses these maximum/minimum value as the feature for detecting the obstacle avoidance behavior. These feature values will be large when a bicycle rapidly slow down. Also, it calculates the minimum values in the difference in the elapsed time between two points at when adjacent local maximum and minimum values are observed.

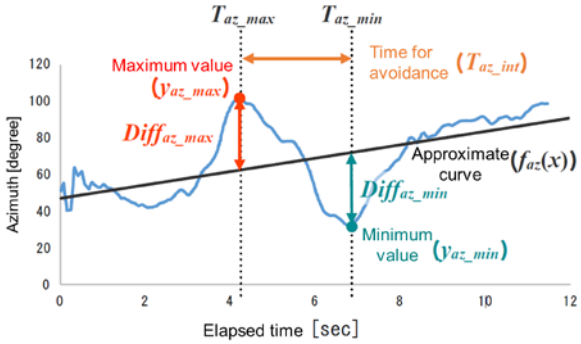


Figure 4: Feature values related on azimuth angle

3.5 Obstacle Detection based on Bicycle Behavior Data

The proposed method estimates the type and size of an obstacle that a bicycle encountered by detecting the specific behaviors for obstacle avoidance from a set of bicycle behavior data. It detects the specific behaviors during the obstacle avoidance, and classifies the obstacles by SVM. Obstacles with different scales can be assumed as classification targets. There are obstacles with small scale such as drainage lids, ones with middles scale such as vehicles parked on road, and ones with large scale such as buses and trucks. In this study, we focus on obstacles with small and large scales that bicycle users may be considered frequently encounter.

3.6 Estimation on Road Surface Condition

We pay attention to the vertical vibration of a bicycle that changes depending on road surface conditions. The proposed method estimates road surface conditions based on the characteristics of the vibration. In this study, road surface conditions are divided into a global condition depending on the quality of the entire road surface such as road materials, and a local condition such as a crack, a hole, and a manhole.

In order to estimate global road surface condition, we focus on the variance of the vertical acceleration, which can capture the difference in the fluctuation of the whole road segment as a feature for estimation. First, the proposed method extracts the segments by points where the fluctuations of the vertical acceleration changes from the bicycle behavior data. Next, it classifies global road surface conditions by using the variance of the vertical acceleration for each extracted segment. Global road surface conditions consists of two categories: “good” and “bad” conditions.

"Good" is defined as the condition when the quality of the entire road surface is smooth and there is almost no vibration. "Bad" is defined as the condition when the quality of the road surface is rough and a bicycle user always feels large vibration.

The proposed method estimates local road surface conditions by detecting the characteristic fluctuations that are locally observed in the targeted road segment. In the case that local road conditions such as cracks and holes due to road surface damage deteriorate, the vibration of local road conditions tend to be more intense than that of global road conditions.

Finally, integrated road surface condition is estimated by using both estimated results of global and local conditions. The proposed method classifies the integrated road condition into three categories: "comfortable", "normal", and "uncomfortable". The details are shown in Table 1.

Table 1: Classification of the integrated road condition

Local road condition	Global road condition	
	good	bad
good	comfortable	normal
bad	uncomfortable	uncomfortable

4 EXPERIMENTS AND DISCUSSIONS

In this section, we describe the experiment to evaluating the effectiveness of the proposed method, and have discussions through the experimental results.

4.1 Experimental Settings

We collected the bicycle behavior data when one subject rode on a bicycle in the real environment. We also recorded the surrounding situations during this data collection by a video camera mounted on the bicycle. There were parked vehicles and drainage lids on the route that the bicycle had passed. We use this data collected from the real environment as an evaluation data set.

We also generated the learning model for obstacle detection from the bicycle behavior data collected in an artificial environment. We prepared an artificial environment that was modeled an actual environment, and collected sensor data of various patterns of bicycle behaviors in that environment. We assumed a straight road, a road with left or right curves as the road shape. We also installed a marker cone that represents a drain lid or a vehicle parked on a road.

We can collect a total of nine patterns of behavior data by combining three types of running courses (straight, left or right curves) and three types of obstacles (no obstacle, a drainage lid, and a road parking vehicles). When doing 10 trials in each behavior pattern for each subject, we can collect the bicycle behavior data for 90 trials for each subject. Since there were three subjects in this experiment, the bicycle behavior data for 270 trials were collected as a training data set for generating a learning model for obstacle detection

4.2 Experimental Results for Obstacle Detection

The obstacle classification result obtained by using the learned model is shown in Table 2. The overall F-value is 0.71. As a whole tendency, we can find many misclassifications between "no obstacle" and "drainage lid", and between "drainage lid" and "parked vehicle". As a result of these misclassifications, it can be inferred that both the precision and recall of "drainage lid" are low.

As a consideration, misclassifications are likely to occur when two behavior data are similar. For example, a behavior in the case that there are no obstacle and a behavior when avoiding a drainage lid gently, a behavior when avoiding a drainage lid and a behavior when gently avoiding a vehicle parked on a road. However, it is essentially difficult to reduce misclassification when two bicycle behaviors in the different situations are similar.

Therefore, we re-consider the index for obstacle classification. This classification index is based on the scale of the obstacles: no obstacle, drainage lid (small scale), and parked vehicle (middle scale). We focus on whether a bicycle can run comfortably that is essential to the purpose of this study. We consider that information about how much avoidance behavior was taken are essential than what kinds of obstacles was avoided as information necessary to recommend a comfortable route.

Table 2: The experimental result of obstacle classification

	Estimated result			Recall
Correct label	No obstacle	Drainage lid (small scale)	Parked vehicle (middle scale)	
No obstacle	27	5	2	0.79
Drainage lid	5	8	0	0.62
Parked vehicle	0	8	20	0.71
Precision	0.84	0.38	0.91	0.71 (F-measure)

Table 3: The classified result by new classification criteria

	Estimated result			Recall
Correct label	Comfortable	Normal	Uncomfortable	
Comfortable	32	5	2	0.82
Normal	0	16	0	1.00
Uncomfortable	0	0	20	1.00
Precision	1.0	0.76	0.91	0.88 (F-measure)

Concretely, we introduce three categories of the comfort level of the road: “comfortable”, “normal”, and “uncomfortable”. We define “comfortable” when there is no obstacle or when a small obstacle such as a drainage lid is gently avoided. “Normal” is defined when a drainage lid is avoided or when a vehicle parked on a road is gently avoided. We define “uncomfortable” when a parked vehicle is suddenly avoided.

We re-labeled the bicycle behavior data based on the new classification index, and conducted the experiment for the classification about the comfort level. The classified result shown in Table 3. The overall F value was 0.88 due to the reduction of misclassifications between similar behaviors in the experiment before changing the classification criteria.

4.3 Experimental Results for Road Surface Conditions

In this section, we describe the experimental results of road condition estimation. First, we calculate the estimation accuracy of the integrated road condition by using only the estimated results of global road condition. The correct label for each divided data is allocated based on the image from the video camera mounted on the bicycle. The accuracy is calculated by 10-fold cross validation. The result is shown in Table 4.

Table 4: The estimated result in the integrated road condition (only using global condition estimation)

	Estimated result		
Correct label	Comfortable	Normal	Recall
Comfortable	38	7	0.84
Normal	18	32	0.64
Precision	0.68	0.82	0.79 (F-measure)

The overall F-measure is 0.79. There was a tendency to misestimate “normal” into “comfortable” road condition. As a result, the precision and recall of “normal” condition is low.

In the case that the estimated result of local road condition as well as that of global road condition are used, the overall F-measure is 0.70. In this case, the integrated road condition is classified into three categories: comfortable, normal and uncomfortable. There were many misclassifications about “uncomfortable”, and overall, the recall and precision were low. Since the estimation accuracy of the local road surface condition is low, it is considered that the estimation result of the integrated road surface condition is low.

5 CONCLUSION AND FUTURE WORKS

The purpose of this study is to recommend comfortable routes for bicycle users, and provide information about the comfort level on the recommended route. This study propose a method for estimating road situations such as the type and scale of obstacles on a road or the unevenness of a road by sensing behaviors when a bicycle is traveling. These information about road situations will highly affect to the comfortability of roads where bicycle users travel. The estimated results are shared among bicycle users based on participatory sensing. Because the penetration rate of smartphones is high, the introduction cost in participatory sensing will be reduced by using a smartphone as a sensing device. A smartphone is mounted on a bicycle, and the bicycle behavior data such as acceleration and angular velocity are automatically collected by using the smartphone.

We focus on detecting obstacles on roads and estimating road surface conditions as representative and important road situations. In this paper, we described a method to detect obstacles on roads and a method to estimate road surface condition. The proposed method for obstacle detection extracts features from the characteristic behavior during obstacle avoidance, and classifies the type and scale of the avoided obstacles based on the extracted features. Also, we pay attention to the vibration of a bicycle that changes depending on road surface conditions, the proposed method estimates road surface condition by using the acceleration data in the vertical direction.

In the evaluation experiment on obstacle detection, under the classification criteria based on the scale of obstacles, the overall F-measure was 0.71 and the accuracy was not high. However, under the new classification criteria based on the scale of avoidance behavior, the overall F-measure was 0.88 and improved the accuracy. We believe that more essential classification has become possible by focusing on the scale of avoidance behavior that is directly related to the comfortability and changing the classification criteria.

In the experiment for estimating road surface condition, the estimation accuracy (F-measure) in the case that we used only the estimated result of global road surface condition was 0.79. If a large amount of data for learning can be collected, it is considered that mis-estimation can be reduced and the overall estimation accuracy can be improved. On the other hand, in the case that we used the estimated results of local and global road surface condition, the estimation accuracy of the integrated road surface condition was 0.70 was and was low.

As one of future works, it is necessary to conduct experiments in actual environments (public roads) with multiple subjects and evaluate the effectiveness of the proposed method. In this experiment, the bicycle behavior data used for the learning phase were collected from multiple subjects, but the test data on public roads were collected from a single subject. It is necessary to evaluate the adaptability of the individual differences in order to make the proposed method more practical. Through the additional experiment in the actual environment, our method will be improved.

A common problem in obstacle detection and road surface condition estimation is to consider a method for collecting learning data efficiently. In the current method, correct label collection is performed manually or visually, and the collection cost is high, and it is difficult to collect a large amount of data as it is.

ACKNOWLEDGEMENT

A part of this work was supported by JSPS KAKENHI Grant Number JP17K00128.

REFERENCES

[1] Ministry of Land, Infrastructure, Transport and Tourism, Bicycle Traffic, Policy Review Result (2014), https://www.mlit.go.jp/common/001085121.pdf, (Accessed 2019-5-31). *(in Japanese)*

[2] Bicycle Life Project, Bicycle Map, http://www.bicyclemap.net/, (Accessed 2019-5-31). *(in Japanese)*

[3] J. Burke, D. Estrin, M. Hansen, A Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava, Participatory Sensing, World-Sensor-Web (WSW '06) at SenSys '06, pp.1-6 (2006).

[4] C. Outram, C. Ratti, and A. Biderman, The Copenhagen Wheel: An Innovative Electric Bicycle System that Harnesses the Power of Real-time Information and Crowd Sourcing, EVER Monaco International Exhibition & Conference on Ecologic Vehicles & Renewable Energies (2010).

[5] N. Shimoyama, H. Ando, K. Yamazaki, T. Ishii, and M. Hirayama, A Study of Bicycle Accident Prevention System by Riding Analysis using Multi-sensor, IPSJ SIG Technical Report (EMB), Vol.36, No.51, pp.1-6 (2015). *(in Japanese)*

[6] Y. Tabata, Y. Kawachi, A. Yamamoto, S. Asada, and S. Kaneda, A System for Sharing Road Obstacle Degree that Provides Obstacle Information to Bicycle Users, IEICE Technical Report (ITS), Vol.113, No.491, pp.1-6 (2014). *(in Japanese)*

[7] A. Zhan, M. Chang, Y. Chen, and A. Terzis, "Accurate Caloric Expenditure of Bicyclists Using Cellphones," Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems, pp.71–84 (2012).

[8] Y. Miyazaki, K. Hishimoto, and T. Kitani, A Method for Estimating Road Surface Damage by Detecting Obstacle Avoidance Motion of Motorcycles, IPSJ SIG Technical Report (ITS), Vol.68, No.6, pp.1-9 (2017). *(in Japanese)*

[9] T. Kamimura, and T. Kitani, A Fundamental Study on Classification of Motorcycle Driving Behavior using SVM, Proceedings of Multimedia, Distributed, Cooperative, and Mobile Symposium (DICOMO 2014), pp.44-52 (2014). *(in Japanese)*

[10] Y. Tabata, A. Yamamoto, Y. Kawachi, S. Asada, and S. Kaneda, Road Hazard Sharing System by Using Mobile Ad Hoc Network for Multiple Cyclists, IEICE Technical Report (KBSE), Vol.113, No.475, pp.73-78 (2014). *(in Japanese)*

[11] H. Saito, K. Sugo, H. Aida, N. Thepvilojanapong, and Y. Tobe, sBike: Acquisition of Person's State Riding a Bicycle with Mobile Sensing for Participatory Sensing, IPSJ Journal, Vol.53, No.2, pp.770-782 (2012). *(in Japanese)*

[12] W. Gu, Y. Liu, Y. Zhou, Z. Zhou, C. J. Spanos, and L. Zhang, Bikesafe: Bicycle Behavior Monitoring via Smartphones, Proceedings of the 2017 ACM International Symposium on Wearable Computers, UbiComp '17, pp. 45-48 (2017).

Session 2:
Multimedia and Groupware
(Chair: Katsuhiko Kaji)

A Communication Support Method with Viewers' POV Capture in 360-degree Internet Live Broadcasting

Yoshia Saito * and Yuka Yamashita*

* Faculty of Software and Information Science, Iwate Prefectural University, Japan
y-saito@iwate-pu.ac.jp

Abstract - In recent years, 360-degree videos have been popular as new entertainment contents. The 360-degree videos are taken by omnidirectional cameras and viewers can freely change their POV (Point of View) to watch all directions of the videos. The POV means directions where the viewers are watching on the 360-degree video. Various well-known web services such as YouTube, Twitter and Facebook support 360-degree Internet live broadcasting which can broadcast 360-degree video to viewers in real time via the Internet. However, there is a critical problem of the 360-degree Internet live broadcasting is that the broadcaster cannot be aware of the viewers' POV. The viewers' POV is important to achieve smooth chat communication between the broadcaster and the viewers in the 360-degree Internet live broadcasting. In this paper, we propose a communication support method which adds viewers' POV capture which means a set of direction and visual information into each viewers' comment.

Keywords: 360-degree video, Internet live broadcasting

1 INTRODUCTION

In recent years, 360-degree videos have been popular as new entertainment contents. In non-360-degree videos taken by conventional cameras, viewers are only able to watch the videos in a specific direction. However, the 360-degree videos are taken by omnidirectional cameras and the viewers can freely change their POV (Point of View) to watch all directions of the videos. The POV means directions where the viewers are watching on the 360-degree video.

Meanwhile, Internet live broadcasting has been also popular as a communication tool for Internet users. In the Internet live broadcasting, a broadcaster can enjoy communication with viewers. The broadcaster reads comments from the viewers and responses to them via the live video.

By combining technologies of 360-degree videos and Internet live broadcasting, 360-degree Internet live broadcasting has been emerged. Since it has features of 360-degree videos and Internet live broadcasting, viewers can freely change their POV and send comments to a broadcaster. The broadcaster can enjoy communication with the viewers by reading their comments. Various well-known web services such as YouTube, Twitter and Facebook support the 360-degree Internet live broadcasting.

However, there is a critical problem of the 360-degree Internet live broadcasting is that the broadcaster cannot be aware of the viewers' POV. In the conventional Internet live broadcasting, the viewers watch a specific direction and send comments to the broadcaster. The broadcaster can easily understand what the viewers talked about by checking a direction of the camera lens. In the 360-degree Internet live broadcasting, it is not enough for the broadcaster to check a direction of the camera lens in order to understand what the viewers talked about because they can watch all directions of the camera. The POV gives expression to object of the current topics for the viewers. If the broadcaster cannot be aware of the viewers' POV, smooth communication between the broadcaster and the viewers is utterly impossible. For example, when a viewer sends a comment about an object which are out of the view sight of the broadcaster, it is difficult for the broadcaster to find the object and understand the comment.

To solve the above problem, we propose a communication support method with viewers' POV Capture which is a function to save direction information and visual information about the viewers' POV so that the broadcaster can understand what the viewers talked about. The contributions of this paper as follows:

- We develop a prototype system of 360-degree Internet broadcasting which has a function of viewers' POV capture. It adds direction information and visual information to comments from viewers.

- We study effect of the direction information and visual information respectively which is added to the comments on communication between a broadcaster and viewers.

The paper is organized as follows. In the next section, we give a detailed description of our proposed method. In section 3, we design and implement a prototype system. We evaluate our proposed method using the prototype system and describe our findings in section 4. Section 5 gives some conclusions and our future work.

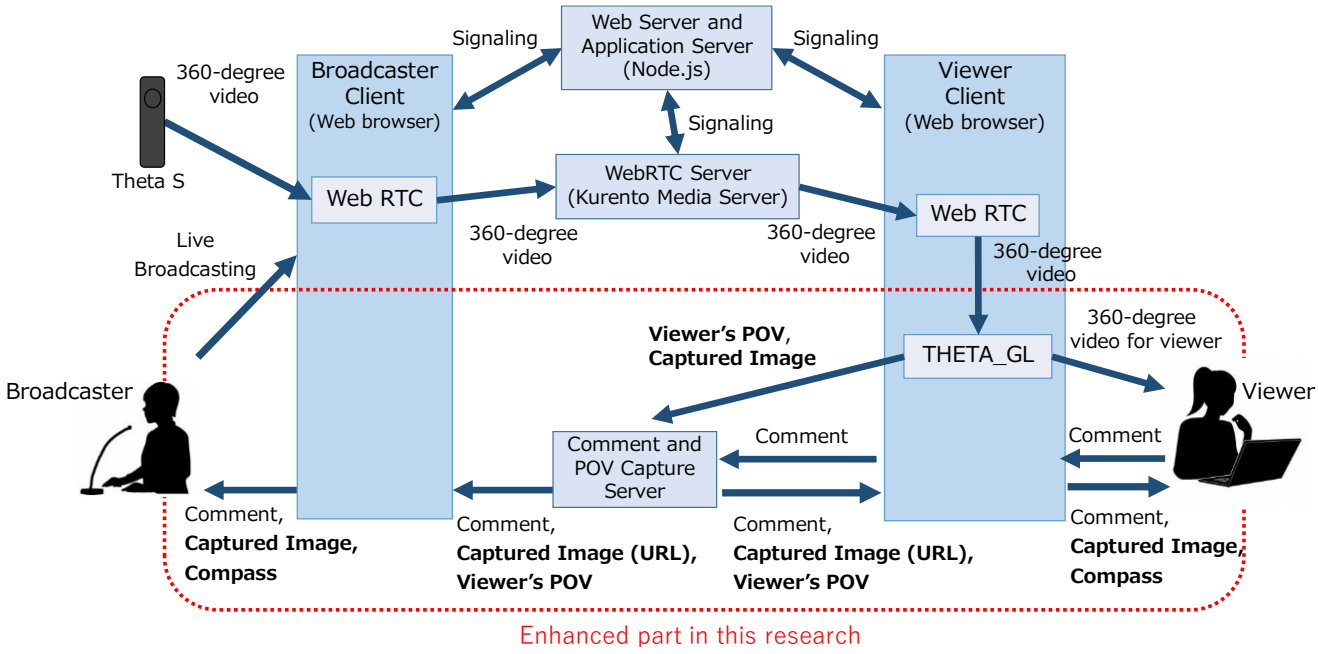


Figure 1: The system architecture of the prototype system

2 PROPOSED METHOD

The proposed method adds direction information and visual information to comments from viewers to reduce time to understand viewer's topics and to achieve smooth communication between the broadcaster and the viewers. The direction information indicates a direction where a viewer watched when his/her comment was sent. The visual information indicates an image which a viewer watched when his/her comment was sent. In this research, we define *viewers' POV capture* as a function which gets a set of the direction information and visual information and adds to the viewers' comments.

In the proposed method, a broadcaster starts 360-degree Internet live broadcasting for viewers via a broadcasting system. The viewers can send text comments to the broadcaster watching the live broadcasting contents in real time. When a viewer sends a comment, a viewer's POV of that time is added to the comment as coordinate information. The viewers' POV capture generates direction information and visual information using the coordinate information of the viewer's POV. The direction information is visualized by a compass metaphor. The visual information is also visualized by a capture of screen image which was watched by the viewer who sent the comment. These direction information and visual information are added to the comment and it is sent to the broadcaster and the viewers.

In the conventional 360-degree live broadcasting, the viewers can communicate with the broadcaster only by text comments. It is difficult for the broadcaster to understand what the viewers are talking about and it is necessary to send comments many times to convey their meaning properly. Since the proposed method provides direction information and visual information, the broadcaster can understand viewer's comment properly. If the viewer sends a comment which includes a demonstrative pronoun, the broadcaster can understand the commented object by reference to the direction information and visual information. It avoids the trouble of having to explain what the comment means and reduces fruitless time in the communication. Furthermore, the viewers also can understand what other viewers are watching and follow the conversation topics. As a result, smooth communication between the broadcaster and the viewers can be realized.

3 PROTOTYPE SYSTEM

We implemented a prototype system to evaluate the effectiveness of the proposed method. Figure 1 shows the system architecture of the prototype system. In the previous work, we have developed a 360-degree Internet live broadcasting system[1] for the prototype system. In this prototype system, we enhanced the 360-degree Internet live broadcasting system. The 360-degree Internet live broadcasting system consists of a broadcaster client, a viewer client, a comment server, a web/application server for web application and a WebRTC server for live video streaming. We used Node.js[2] for the web/application server and Kurento media server[3] for the WebRTC server. The broadcaster and viewer client can be used on web browsers as web application.

We use Theta S[4] as an omnidirectional camera. The 360-degree video taken by the camera is rendered on a sphere object using THETA_GL[5] so that the viewers can watch the video changing their POV. The viewer's POV is represented by spherical coordinates (r, θ, φ). r denotes a radius of the sphere, θ and φ denotes deflection angles.

The broadcaster can start 360-degree live broadcasting using the broadcaster client. The 360-degree video are sent to the viewers via the WebRTC server and it is shown on the

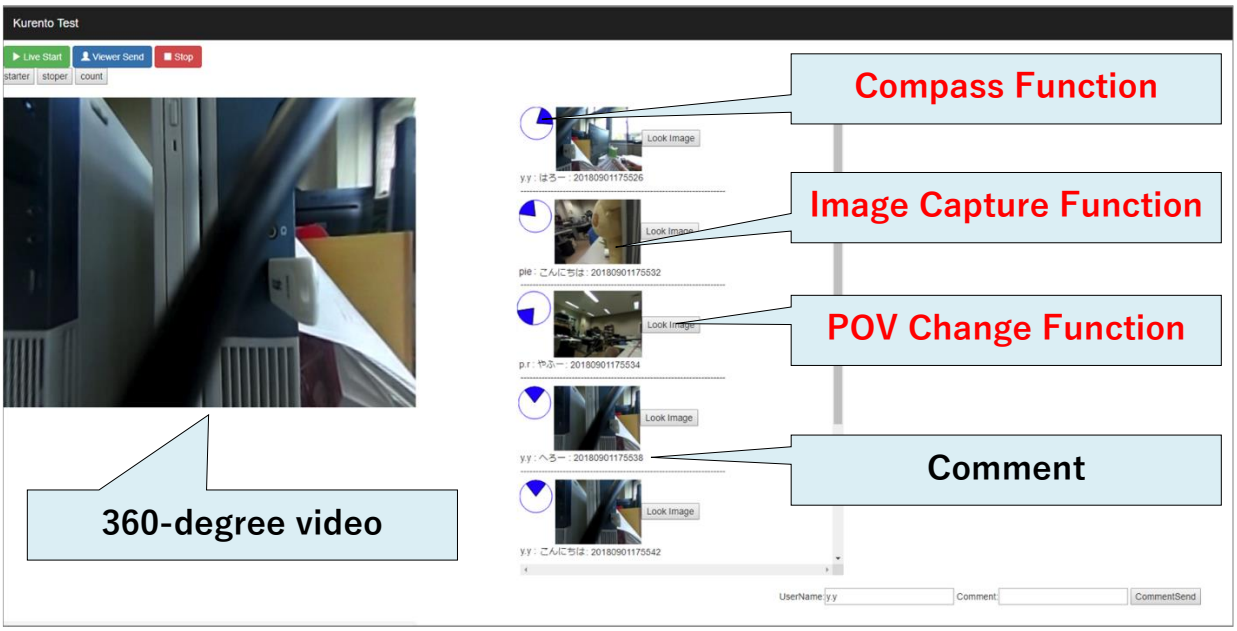


Figure 2: User interface for the viewer client

viewer client. The viewers can freely change their POV to watch all directions of the videos same as typical 360-degree video services. The viewers also can send text comments to the broadcaster via the comment server. We enhanced the comment server to realize the viewers' POV capture. We implemented several functions which are compass function, image capture function and POV change function as key functions of the prototype system.

3.1 Compass Function

The compass function converts a spherical coordinate of the viewer's POV into a two dimensional direction information. The direction information is visualized by a compass metaphor. As a drawing method of the compass, the direction of the viewer's POV is drawn by a colored sectorial figure on a circle. We adjust the size of the sectorial figure so that it can match with a display range of the video on the viewer's client.

The standard direction is the front of the omnidirectional camera and it indicates the upper direction of the compass. The sectorial figure of the compass moves in accordance with horizontal movement of the viewer's POV. The broadcaster can understand the direction of the viewer's comment by setting up the direction of the front of the omnidirectional camera to coincide with that of the broadcaster.

The compass is shown on the upper part of each comment. We set the radius of the compass to 20px in order to be able to view a comment log without being disturbed by the compass. Each set of a compass and comment is divided by dotted separators to improve easiness to read the comments.

3.2 Image Capture Function

The Image capture function provides a screen image which is displayed when the viewer commented. The viewer client generates a captured screen image from the video when the viewer commented. The screen image is sent to the POV capture server with the comment and the viewer's POV. The POV capture server stores the image data on the local filesystem. Then, a URL address of the image data is sent to all clients with the comment and the viewer's POV. The clients read the image data from the received URL and display it with the comment. The screen image is shown on the upper part of each comment.

3.3 POV Change Function

The POV change function moves the viewer's POV according to the POV of the comment. If the viewers want to confirm an object which other viewers watch, there is a possibility that direction information and visual information are not enough to follow the conversation topics. On the viewer client, the button of the POV change is displayed on the side of the screen image with the comment. The viewers can push the button when they want to changes their POV to the POV of the comments. For the evaluation, we save the number of click counts of the POV change button on the POV capture server.

3.4 User Interface for Clients

Figure 2 shows the user interface for the clients. Since the difference of the user interface between the broadcaster client and the viewer client is only whether the 360-degree

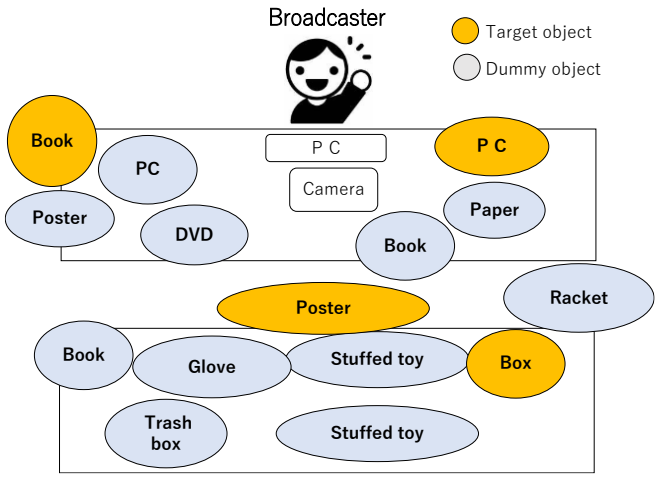


Figure 3: the layout of the objects in the broadcaster's room

video is shown or not, the figure shows that of the viewer client. The compass function, the image capture function and the POV change function are displayed as a compass, a screen image and a button respectively. The comments are sorted from newest to oldest and the latest five comments are displayed.

4 EVALUATION

We conducted an experiment to evaluate effectiveness of the communication support method with viewers' POV capture in 360-degree Internet live broadcasting.

4.1 Experiment Environment

We conducted the evaluation experiment two times. Participants of the experiment were 8 students in our University. In each experiment, 4 students participated and included 1 broadcaster and 3 viewers. We compared the conventional method which used only text comments for the communication with the prototype system as our proposed method in terms of time when the broadcaster could understand the commented object. The broadcaster performed 360-degree Internet live broadcasting of the conventional method and the proposed method in order. Each broadcasting was performed for 13 minutes. We changed the order of the broadcasting to keep fairness in the evaluation.

The experiment task was same in the conventional method and the proposed method. For the first 3 minutes of the broadcasting, the broadcaster performed freely broadcasting in order to get used to operation of the system. The viewers watched the broadcasting and sent comments freely. One of the viewers sent a comment every 4 minutes according to the Instructions. The comments were a question included a demonstrative pronoun. If the questioner felt that the broadcaster did not understand the object of the comment, he/she sent several additional comments and

Table 1: The questionnaire after each broadcasting

Subject	Question	Method
Broadcaster	Did you understand an object which a viewer indicated soon after reading the comment?	Five-grade evaluation
Broadcaster and Viewers	Was this system easy to use?	Five-grade evaluation
	Do you have any good/bad point of the broadcasting?	Free descriptive answer
Viewers	Did you understand what object other viewer indicated soon after reading the comment?	Five-grade evaluation
	Did you understand the object's location which other viewer indicated soon after reading the comment?	Five-grade evaluation

Table 2: The questionnaire for functional evaluation

Subject	Function	Method
Broadcaster and Viewers	Comment	Five-grade evaluation
	Compass	Five-grade evaluation
	Image capture	Five-grade evaluation
Viewers	POV change	Five-grade evaluation

guided the broadcaster to the target object. The broadcaster read aloud all comments which had <Instruction> at the head and raise his/her hand when understand the target object. We measured the time period between from the beginning of reading the comment to the end of raising the broadcaster's hand. In the experiment, the broadcaster and the viewers divided into different 2 rooms. In the broadcaster's room, there were several objects for the instruction comment. Figure 3 shows the layout of the objects.

The flow of the evaluation experiment was as follows. At first, the broadcaster performed broadcasting using the conventional method. The broadcaster and the viewers filled in a questionnaire as shown in Table 1. For the broadcaster, understandability of the object which a viewer indicated was evaluated on a scale of one to five. For the broadcaster and the viewers, usability of the system was evaluated on a scale of one to five and good/bad points of the broadcasting were evaluated by using a free descriptive answer text. For the viewers, understandability of the object which other viewer indicated was evaluated on a scale of one to five.

Next, the broadcaster performed broadcasting using the proposed method. After that, the broadcaster and the viewers filled in the same previously mentioned questionnaire as shown in Table 1. Furthermore, the broadcaster ad the viewers filled in another questionnaire for functional evaluation as shown in Table 2. In the questionnaire, functions of the comment, the compass, the image capture and the POV change were evaluated on a scale of one to five. We compared the proposed method with the conventional method using these results.

4.2 Evaluation Results of Broadcaster

Figure 4 shows the result of time period until the broadcaster understood the object which is indicated by the instruction comments. The proposed method reduced the average of the time period from 38.25 seconds to 9.5 seconds in comparison with the conventional method. This result shows the proposed method can reduce fruitless time for the communication between the broadcaster and the viewers.

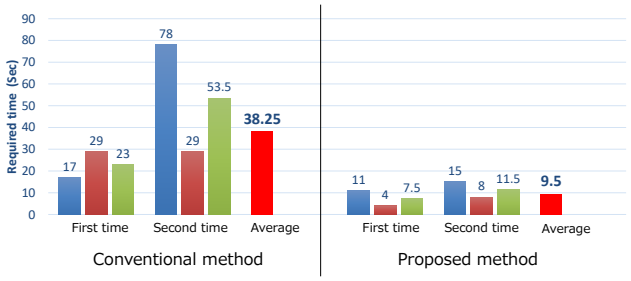


Figure 4: Time period until the broadcaster understood the object which is indicated by the instruction comments

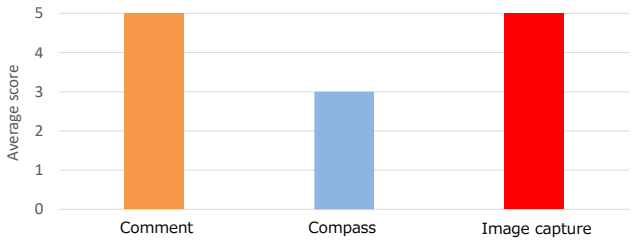


Figure 5: The results of functional evaluation for the broadcaster

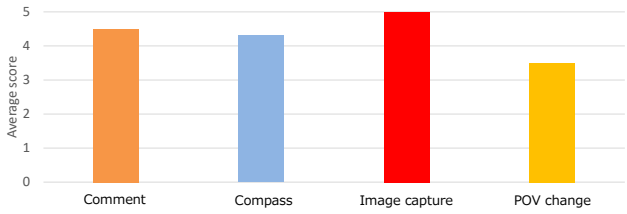


Figure 6: The results of functional evaluation for the viewers

Figure 5 shows the results of functional evaluation. The average scores of the comment function, the compass function and the image capture function were 5.0, 3.0 and 5.0 respectively. From the result, the image capture function was most effective and the visual information contributed to improve understandability of the communication. On the other hand, the compass function got low score in the evaluation. This result shows the direction information plays a supplementary role of the visual information.

4.3 Evaluation Results of Viewers

Figure 6 shows the results of functional evaluation. The average scores of the comment function, the compass function, the image capture function and the POV change function were 4.5, 4.3, 5.0 and 3.5 respectively. From the result, the image capture function was most effective as same as the result of the broadcaster and the visual information also contributed to improve understandability of the communication for the viewers. About the POV change function, it was used only 1 time in each broadcasting and it hardly had any chance to be used. If the viewers wanted a direction where other viewer watched, we found there were no demand of the functionality such as the POV change.

5 CONCLUSION

In this paper, we proposed a communication support method with viewers' POV capture in 360-degree Internet live broadcasting to solve a problem which was that it was difficult to archive smooth communication between the broadcaster and the viewers because the broadcaster could not understand the viewers' POV. We implemented a prototype system and evaluated the effectiveness of the proposed method comparing with the conventional method. From the evaluation results, the proposed method, for the broadcaster, reduced time to understand viewer's topics and understandability of the comments from the viewers. The proposed method also improved understandability of the comments among the viewers. In the functions of the proposed method, the image capture function which provided the visual information was most effective and it was an important factor to support communication between the broadcaster and the viewers.

REFERENCES

[1] Masaya Takada, Dai Nishioka and Yoshia Saito: Proposal of a Spherical Heat Map in 360-Degree Internet Live Broadcasting Using Viewers' POV, The 2nd International WOrkshop on Mobile Ubiquitous Systems, Infrastructures, Communications and Applications (MUSICAL2019), pp.596-600 (2019).
[2] Node.js, <https://nodejs.org/>.
[3] Kurento, <http://www.kurento.org/>.
[4] Theta S, < https://theta360.com/en/about/theta/s.html>.
[5] Theta_GL, <https://github.com/mganeko/THETA_GL>.

A Video Processing System to Stabilize Frame Rates on Trust-oriented Internet Live Video Distributions

Tomoki Yoshihisa[*], Satoru Matsumoto[*], Tomoya Kawakami[**], and Yuuichi Teranishi[***,*]

[*]Cybermedia Center, Osaka University, Japan
[**]Nara Institute of Science and Technology, Japan
[***]National Institute of Information and Communications Technology, Japan
yoshihisa@cmc.osaka-u.ac.jp

Abstract - Due to the recent popularization of Internet live video distributions, Internet live video distributors such as YouTubers have attracted great attention. Some of them hide the image regions related to personal information, e.g., their faces or current locations, so as not to encounter public concerns such as threats or attacks to them. In the cases that the video processing time to hide such image regions is long, the frame rate fluctuates and unstable frame rates annoy the viewers. Hence, in this paper, we propose a video processing system to stabilize frame rates on trust-oriented Internet live video distributions [1]. Our proposed system adopts the approach that changes the video process when the video processing time is going to exceed the interval of the video frame. Our evaluation results revealed that our proposed system can stabilize the frame rate for trust-oriented Internet live video distributions.

Keywords: Broadcasting, YouTubers, Continuous Media, Data Streaming, Video-on-Demand

1 INTRODUCTION

Due to the recent popularization of Internet live video distributions, Internet live video distributors have attracted great attention. Most of Internet live distributors distribute the videos shooting themselves by cameras. For example, video distributors on YouTube are called YouTuber and 1 million of YouTubers distribute the videos shooting themselves. They often distribute live videos. Some of them hide image regions related to personal information, e.g., their faces or current locations, so as not to encounter public concerns such as threats or attacks by the viewers. It has a large possibility that public concerns do not occur if there is a trust between Internet live video distributors and the viewers because the trust construct their social relations. Therefore, trust-oriented Internet live video distributions proposed in [1] can cause safer and wider used Internet live video distributions.

In the cases that the distributors shoot themselves, the most sensitive personal information is their faces. To avoid public concerns, some of them change or hide their faces by adding video effects [2]. Such video effects include some processes for detecting their faces, creating mask images, and drawing the mask images to their faces, and have a higher computational load compared with a simple process. In the cases that the video processing time is longer than the interval of the video frames, the time to draw the processed image for a video frame delays and the frame rate decreases. The video processing times depend on the complexity of the processes and the images, and the delays cause unstable frame rates. Unstable frame rates annoy the viewers. Therefore, stable frame rates are required for Internet live video distributions.

Various techniques to reduce video processing time have been proposed. Some of them give an upper limit on the video processing time and cancel the process when the processing time reaches to the upper limit. In the cases that the video process is changing or hiding the distributor's face, the face appears in the video when the processing time reaches to the upper limit because the process is cancelled. This is not a trust-oriented Internet live video distribution since the video exposes the distributor's personal information even when there is no trust between the distributor and the viewers. In trust-oriented Internet live video distributions, distributors' personal information should be hidden when there is no trust. However, existing video processing time reduction techniques do not consider trust and cannot realize trust-oriented Internet live video distributions.

In this paper, we propose a video processing system to stabilize framerates on trust-oriented Internet live video distributions. In trust-oriented Internet live video distributions, the distribution situation is classified into two situations. One is the situation that there is no trust between the distributor and the viewers. The other one is the situation that there is trust. In the former situation, our proposed system always hide the distributor's personal information. To achieve this, our proposed system changes the video processes when the video processing time is going to exceed the interval of the video frame to a simple process. For example, in the cases that the time is close to draw the image for the next frame while the processing computer executes the process to detect the distributor's face in the video, the computer cancels the face detection process and starts executing the process to blur the whole region of the image. In addition, we implement our proposed system and evaluate its performance. The main novelty of the paper is the stabilization of the framerates considering trust on Internet live video distribution. Although some previous researches try to stabilize framerates, they do not consider trust and wholly give up to execute video processing when the processing time reaches to the upper limit. Our proposed system change the process considering to manage the trust. The contributions of this research are; 1) the proposition of a

[1] The work was supported by a Grants-in-Aid for Scientific Research (C) numbered JP18K11316, and by I-O DATA Found.

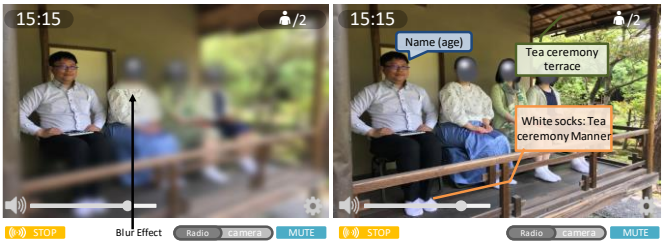


Figure 1: Live video images for an un-trusty situation (left) and a trusty situation (right).

video processing system to stabilize framerates on trust-oriented Internet live video distribution, 2) the design and the development of the system, 3) the evaluation of the system.

The result of the paper is organized as follows. We introduce related work in Section 2. We explain the trust-oriented Internet live video distributions in Section 3 and our proposed video processing system to stabilize frame rates in Section 4. We show some evaluation results and discuss about them in Section 5. Finally, we will conclude the paper in Section 6.

2 RELATED WORK

Many systems have been proposed for distributed stream processing (DSP) in a peer-to-peer (P2P), cloud, edge, or fog computing model [3-13]. Some of the proposed systems assume or are implemented by open-source stream processing platforms such as Apache Hadoop [14], Storm [15], Flink [16], and Spark Streaming [17]. Lopez et al. carried out two experiments concerning threats detection on network traffic to evaluate the throughput efficiency and the resilience to node failures for Storm, Flink, and Spark Streaming [18]. The results show that the performance of native stream processing systems, Storm and Flink, is up to 15 times higher than the micro-batch processing system, Spark Streaming. On the other hand, Spark Streaming was robust to node failures and provided recovery without losses.

Some of the DSP systems are designed for real-time processing and can be applied for live video streaming. In [5], the proposed scheme determines the evaluation order for the conditional expressions in continuous queries to reduce the processing time. Its evaluation shows that the proposed scheme can reduce the maximum number of communication hops and the average amount of communication traffic in IoT environments. Ning et al. proposed Mobile Storm as a distributed real-time stream processing system for mobile cloud [7]. Without offloading computation to remote servers, Mobile Storm processes real-time streaming data using a cluster of mobile devices in a local network. Mobile Storm was implemented on Android phones, and a video stream processing application was developed to evaluate its performance. The results show that Mobile Storm is capable of handling video streams of various frame rates and resolutions in real-time. Choi et al. proposed DART as a fast and lightweight stream processing framework for the IoT [9]. In DART, a logical group of data sources, namely, a Cloud of Things (CoT) is composed to process the data streams more efficiently in a fully distributed fashion. DART aims to overcome both server-based and edge-only-based methods by grouping IoT devices as a CoT. RIDE was proposed to

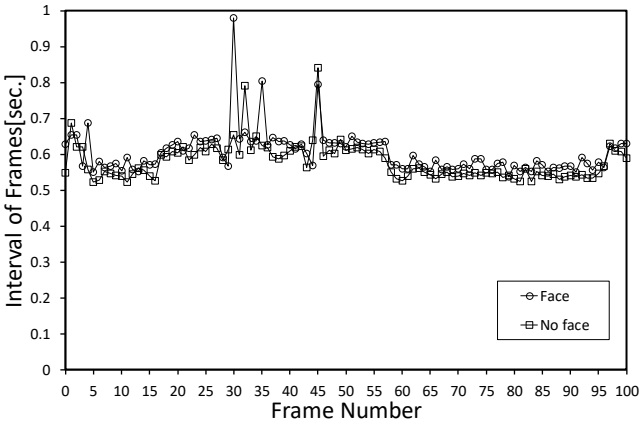


Figure 2: Example intervals of frames

process real-time massive image stream on distributed environment efficiently [10]. RIDE consists of four layers: application, master, buffer, and worker layers. To minimize the communication overhead between the tasks on distributed nodes, coarse-grained parallelism is achieved by allocating partitions of streams to worker nodes in RIDE. In addition, fine-grained parallelism is achieved by parallel processing of task on each worker node. Yang et al. focused on distributed fault-tolerant processing (DFP) method and proposed a distributed image-retrieval method designed for cloud-computing based multi-camera system in smart city [11]. Through the combination of the cloud storage technology, data encryption, and data retrieval technology, efficient integration and management of multi-camera resources are achieved. In [12], processing and bandwidth issues for a typical video analytics application were investigated to help understand placement decision of methods between edge and cloud. The authors in [12] also say that there are further considerations than made in [12], such as privacy, central sharing, edge device maintenance, and processing and bandwidth costs.

The existing techniques mentioned above can reduce the stream processing time by efficiently distributing the computational and communication loads to the processing nodes. However, those techniques do not give an upper limit for the processing time and there is a large possibility that the frame rate fluctuates. This paper is positioned to the consideration of the situations such as allowed processing time and adaptive task modification on the trust-oriented Internet live video distribution.

3 TRUST-ORIENTED INTERNET LIVE VIDEO DISTRIBUTIONS

In this section, we briefly explain trust-oriented Internet live video distributions proposed in [1].

In trust-oriented Internet live video distributions, the distribution situation is classified into two situations. One is the situation that there is no trust between the distributor and the viewers and is called the *un-trusty* situations. The other one is the situation that there is trust and is called the *trusty* situation. The situation that an Internet live video distribution belongs to depends on various factors. For example, in the cases that the viewers of an Internet live video distribution is limited to only the friends of the distributor, the situation is a

trusty situation. In the cases that the viewers are the stranger of the distributor, the situation is an un-trusty situation. In a simple system, the un-trusty/trusty situation that the current situation belongs to is selected by the distributor manually. Automatic selection is also possible by using the information about the viewers, the locations, and so on. Figure 1 shows example images of live video in an un-trusty situation and a trusty situation.

The trust-oriented Internet live video distributions have three policies, “close-information”, “limit-information”, and “expose-information” policies. In the un-trusty situation, the distributors may use the close-information policy. In this policy, the processing computer executes the processes so as to close personal information. The distributor can safety distribute the shot live video since the distributor’s personal information is closed. However, the possibility to be able to construct trust decreases since the viewers cannot get the personal information about the distributor so any more. In the limit-information policy, the processing computer executes the processes accepted by the distributor. The distributors select this policy aiming to keep the trust. In the trust situation, the distributors may use the expose-information policy. In this policy, the processing computer executes the processes so as to expose personal information. The possibility to be able to construct trust increases since the viewers can get the personal information about the distributor. However, the live video distribution is unsafe.

In the trust-oriented Internet live video distribution, the processing computer executes video processes based on the policy selected by the distributor (un-trusty or trusty). However, conventional trust-oriented Internet live video distribution systems do not consider the processing time and the frame rate usually fluctuate.

4 PROPOSED SYSTEM

We explain our proposed video processing system to stabilize frame rates in this section. We explain our target problem first and our approach to solve the problem after that.

4.1 Target Problem

As described in Section 1, long video processing times cause unstable frame rates and this annoy the viewers. For example, we show the intervals of the frames in Figure 2. The horizontal axis is the frame number. In this example, the processing computer executes the processes to detect faces and to blur the detected region. “Face” indicates the interval of the frames when there is a distributor’s face in the live video. “No face” indicates that when there is no distributor’s face. As shown in this example, the interval of the frames fluctuate and have a range of approximately 200 [msec.] in this case. Such a large change of the interval has a large possibility to annoy the viewers. Therefore, in this paper, we propose a video processing system aiming to stable frame rates.

In the cases that the processing time is shorter than the interval of the frames, the system can control the time to draw the image for the next frame by waiting for some time. Otherwise, the time to draw the image for the next frame delays and the frame rate changes. We propose a video processing system to stabilize frame rates even when the processing time is long.

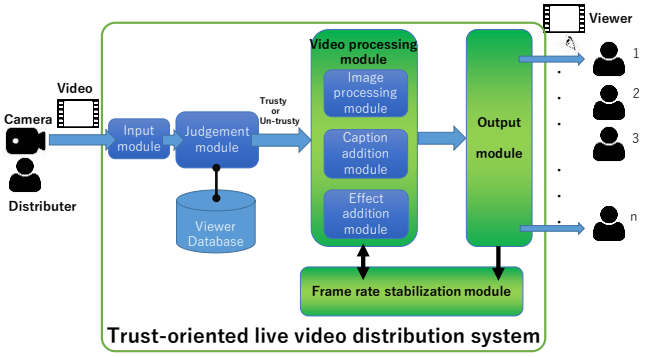


Figure 3: The architecture of our proposed system

4.2 Our Approach

To stabilize frame rates for trust-oriented Internet live video distributions, we adopt two approaches.

The first one is the reduction of the video processing time. As explained in the previous subsection, the system can stabilize the frame rate if the processing time is shorter than the interval of the frames. Therefore, shorter video processing times than the interval of the frames enables stable frame rates. Various techniques to reduce video processing time has been proposed. However, they do not focus on the trust-oriented Internet live distributions and some public concerns can occur even when the system adopts these techniques. The most sensitive personal information is their faces. Therefore, we focus on hiding the distributor’s faces and propose a video processing time reduction method for hiding the face in the trust-oriented Internet live video distributions.

The other one is the change of the video process to a simple video process. As explained in Section 1, most of video effects require a higher computational load compared with a simple process. Therefore, our proposed method changes the video process when the video processing time is going to exceed the interval of the video frames to a video effect that the processing computer can execute the process with a shorter processing time. An example of a simple video process is blurring the whole region of the image. Since the processing time of a simple video process is relatively short compared with complicated video processes, the system can finish the process before the time to draw the image for the next frame and thus can stabilize the frame rate.

4.3 Proposed System Architecture

Figure 3 shows the architecture of our proposed system. In our proposed architecture, the distributors distribute their shot live videos to the viewers using the trust-oriented live video distribution system. The situation judgement module judges whether the current situation is an un-trusty or an trusty situation using the viewer database and other information. The video processing module executes the designated video processes. A part of the viewers login to the system before watching the live videos. The details of these modules are inscribed in [1].

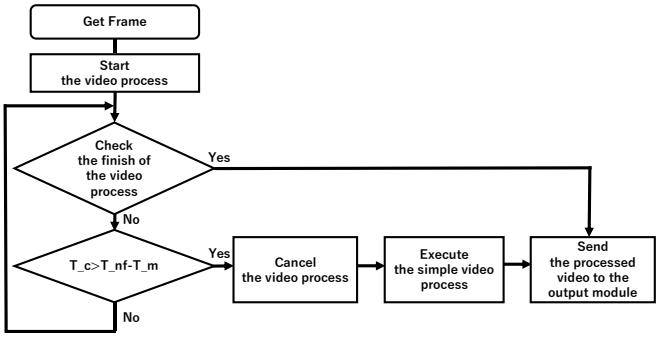


Figure 4: The flowchart for the frame rate stabilization module

The frame rate stabilization module is newly added to the system in our proposed system. The frame rate stabilization module manages the video processing module and checks the frame rates of the output module. The output module is the system module to distribute the live videos to the viewers. When the frame rate is going to fluctuate, the frame rate stabilization module changes the video process executed in the video processing module to a simple process and try to reduce the processing time. The processed video data are transferred to the output module and are distributed to the viewers.

4.4 A Method to Change Video Process

In this subsection, we will explain our proposed method to change video process to stabilize the frame rates.

4.4.1. Timing to Change Video Process

In our proposed method, the frame rate stabilization module changes the video process when the processing time is going to exceed the interval of the frame to a simple process that the video processing module can execute the process with a shorter processing time. Even when the video process is changed to the simple one, the video processing time arises and it takes some time. Therefore, our proposed method gives a margin time for the simple video process. Let T_m denote the margin time. When the current time satisfies the following inequality, the frame rate stabilization module changes the video process to the simple one.

$$T_c > T_{nf} - T_m \quad (2)$$

Here, T_c is the current time and T_{nf} is the time to draw the next frame image. That is, the value of $T_{nf} - T_m$ is the *deadline* of the original video process. Our proposed method decides T_m enough to finish the simple video process within the margin time. For example, in our brief experiment, the maximum processing time to blur the whole region in a video processing system is 129 [msec.]. In this case, we can predict the sufficient time to finish the simple video process although the time fluctuates. In this case, the value of T_m larger than 129 [msec.] is sufficient.

In the cases that the right formula of Equation (2) is very close to the left, the video process can frequently change. This can cause the annoyance of the views. To determine T_m, it is better to consider such an influence of the viewers.

4.4.2. Video Process to Hide Face

Figure 4 shows the flowchart for the frame rate stabilization module. When the module gets a video frame image from the camera, it starts the video processing to the image in other thread. To manage the processes, the video processes are executed in parallel and in other thread from the frame rate stabilization module. After that, the module continuously checks whether the process finishes and whether the above inequality is satisfied or not. When the process finishes, the module transfers the processed image for the video frame to the output module. When the above inequality is satisfied, the module cancels the video process and start executing the simple video process. When the simple video process finishes, it transfers the processes image for the video frame to the output module.

5 EVALUATION

In this section, to evaluate the performance of our proposed system, we show some evaluation results about the intervals of the frames and the video processing.

5.1 Evaluation Setting

We implemented the video processing system that detects the faces in the shot video images and blurs the detected region. In the cases that the time to finish the video process is longer than the deadline (Equation (2)), the system cancels the processing and changes it to the simple process that blurs the whole region of the image. The system has a function to distribute the live video, but the function is not directly related to this research and we do not focus on this function.

To show the effectiveness of our proposed method, we show the intervals of the frames. We set the frame rate of the video to f [fps] in this evaluation. In our proposed method, when the current time exceeds the deadline, the processing computer changes the video process to the simple one. Therefore, in the cases that the interval of the frame is shorter than the $1/f$ [msec.], our proposed method can achieve a constant frame rate by waiting for drawing the next frame image.

We use the ratio of changing process in this evaluation. The ratio means the ratio that the video process is changed to the simple one and is the number of the simple video process divided by the number of the all video process. A larger value means that a more number of the video frames is wholly blurred (the whole region is blurred). A smaller value is better for the trust-oriented Internet live distributions since just for the region related to the personal information is blurred.

5.2 Evaluation Environment

For the evaluation, we use our developed video processing system. We developed the system using the Visual Studio 2017 and the system uses OpenCV 4.1.0 to get the images from the camera and to detect the faces in the images. The processing computer is a laptop computer (CPU: Core i7@2.4GHz, Memory: 8GB). We use the camera equipped on the laptop and get 640x480 RGB (32bits) images. For the face

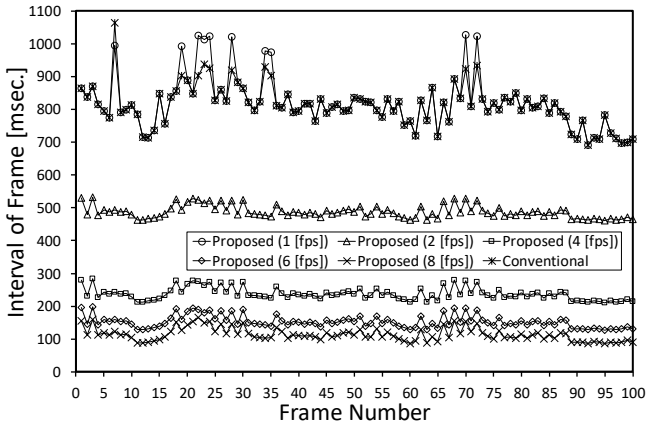


Figure 5: The frame number and the intervals of the frames

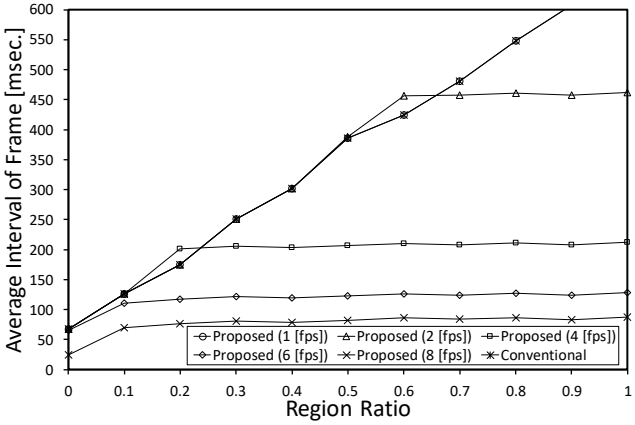


Figure 6: The average interval of the frames under the different region sizes for the processes

detection, we use the detectMultiScale function implemented in the OpenCV.

5.3 Evaluation Results

In this subsection, we show some evaluation results. First, we show the intervals of the frames and evaluate the effectiveness of our two approaches. After that, we show the performances changing the region sizes for the processes to evaluate our approach to change the region size to hide the face faster. Then, we show the performances changing some parameters for the face detection. We evaluate our proposed system under the situations that the video records just one faces although our proposed system can detect multiple faces to make the evaluation results easily understandable. In the situations that the video records multiple faces, their positions also influence the results.

5.3.1 Interval of Frame

The target problem of this research is the stabilization of the frame rate. This means that the intervals of the frames are more constant. Therefore, we measured the intervals of the frames.

Figure 5 shows the intervals of the frames. The horizontal axis is the frame number and the vertical axis is the intervals of the frames. In the figure, "Proposed (*f* [fps])" indicates the intervals of the frames under our proposed method when the frame rate is set to *f*. "Conventional" indicates the intervals of the frames under conventional methods, i.e., without our proposed method. The result under the conventional method does not depend on the frame rate since the method does not consider the frame rate. The margin time is 100 [msec.]

From this result, we can see that our proposed method achieves a shorter interval than the inverse value of the frame rate in many cases. This means that our proposed method gives stable frame rates. However, the intervals of the frames are sometimes longer in the cases that the video processing time is longer than the predicted margin time. The conventional method can achieve almost 1 [fps]. In the cases that the frame rate is 1 [fps], some intervals of the frames under our proposed method is the same as that under the conventional method because the video process finishes before the deadline. However, in the cases that the video process before changing it finishes earlier than the simple video process, the intervals of the frames under our proposed method is longer than that under the conventional method. For example, in the case of 1 [fps], the video processing times of 6 frames in 100 frames exceed the time to draw the image for the next frame. In the case of 4 [fps], the video processing times of 19 frames in 100 frames exceed the time to draw the image for the next frame. We can reduce the number of the frames in that the processing time exceeds the time to draw the image for the next frame by increasing the margin time.

The ratio of changing the process increases in proportional to the frame rate since a larger frame rate gives an earlier deadline. For example, in the cases of 1 [fps], 10 video processes are changed to the simple video processes in 100 frames and this means the ratio of changing the process is 10/100=0.1. In the cases of 4 [fps], all video processes are changed to the simple video processes and the ratio is 1.0. Therefore, Our proposed method gives stable frame rates but the ratio of changing process increases.

In the following evaluation results, we use the average interval of frame.

5.3.2 Region Size for Process

A smaller region size for the process gives a shorter video processing time because the data amount for executing the video process decreases. Hence, changing the region size for the process, we measured the average interval of the frames and the ratio of changing the processes.

Figure 6 shows the average intervals of the frames under the different region sizes for the processes. The horizontal axis is the region ratio. The region ratio is the ratio of the region for the process in the whole region and given by the number of the pixels in the region for the process divided by that in the whole region. The vertical axis is the average intervals of the frames. The margin time is 100 [msec.]. The legends are similar to the previous results.

In our proposed method, when the region ratio is small, the average interval of the frames increases as the region ratio increases since the data amount for executing the video process increases. When the region ratio is large, our proposed method gives almost constant average interval of the frames since the video process is changed to the simple video process when the video processing time reaches to the deadline. The video processing time earlier reaches to the

deadline as the frame rate increases. The average interval of the frames under the conventional method increases in proportional to the region ratio, since the method does not consider the deadline. For example, the average interval of the frames in the cases of 4 [fps] is approximately 200 [msec.] This is smaller than 250 [msec.] (the interval of the frames under 4 [fps]) and our proposed method can give a stable frame rate. On the other hand, in the conventional method, the region ratio should be less than 0.23 to achieve the frame rate of 4 [fps].

6 CONCLUSION

In this paper, we propose a video processing system to stabilize framerates on trust-oriented Internet live video distributions. Our proposed system changes the image region for video process and also changes the video process when the video processing time is going to exceed the interval of the video frame. We implemented our proposed system. Our evaluation results revealed that our proposed system can stabilize the frame rate for trust-oriented Internet live video distributions. In the future, we will further reduce the video processing time and investigate the frame rates in practical Internet live video distributions.

REFERENCES

[1] T. Yoshihisa, S. Matsumoto, T. Kawakami, and Y. Teranishi, Trust-oriented Live Video Distribution Architecture, in Proc. IEEE Int'l Conf. on Computers, Software and Applications, to appear (2019).

[2] S. Matsumoto, T. Yoshihisa, T. Kawakami, and Y. Teranishi: A Distributed Multi-Viewpoint Internet Live Broadcasting System with Video Effects, in Proc. International Workshop on Informatics (IWIN'18), pp. 83-88 (2018).

[3] X. Zhao, H. Ma, H. Zhang, Y. Tang, and Y. Kou, HVPI: Extending Hadoop to Support Video Analytic Applications, in Proceedings of the 8th IEEE International Conference on Cloud Computing (CLOUD 2015), pp. 789-796 (2015).

[4] W. Kou, H. Li, and K. Zhou, Turning Video Resource Management into Cloud Computing, Future Internet 8(3), 35, 10 pages (2016).

[5] T. Yoshihisa, T. Hara, A Low-Load Stream Processing Scheme for IoT Environments, in Proceedings of the 2016 IEEE International Conference on Big Data (Big Data 2016), pp. 263-272 (2016).

[6] N. Chen, Y. Chen, Y. You, H. Ling, P. Liang, and R. Zimmermann, Dynamic Urban Surveillance Video Stream Processing Using Fog Computing, in Proceedings of the 2016 IEEE Second International Conference on Multimedia Big Data (BigMM 2016), pp. 105-112 (2016).

[7] Q. Ning, C.-A. Chen, R. Stoleru, and C. Chen, Mobile Storm: Distributed Real-Time Stream Processing for Mobile Clouds, in Proceedings of the 4th IEEE International Conference on Cloud Networking (CloudNet 2015) (2015).

[8] T. Li, J. Tang, and J. Xu, A Predictive Scheduling Framework for Fast and Distributed Stream Data Processing, in Proceedings of the 2015 IEEE International Conference on Big Data (Big Data 2015), pp. 333-338 (2015).

[9] J.-H. Choi, J. Park, H. D. Park, and O.-G. Min, DART: Fast and Efficient Distributed Stream Processing Framework for Internet of Things, ETRI Journal, Vol. 39, No. 2, pp. 202-212 (2017).

[10] Y.-K. Kim, Y. Kim, and C.-S. Jeong, RIDE: Real-Time Massive Image Processing Platform on Distributed Environment, EURASIP Journal on Image and Video Processing, 13 pages (2018).

[11] J. Yang, B. Jiang, and H. Song, A Distributed Image-Retrieval Method in Multi-Camera System of Smart City Based on Cloud Computing, Future Generation Computer Systems, Vol. 81, pp. 244-251 (2018).

[12] L. O'Gorman and X. Wang, Balancing Video Analytics Processing and Bandwidth for Edge-Cloud Networks, in Proceedings of the 24th International Conference on Pattern Recognition (ICPR 2018), pp. 2618-2623 (2018).

[13] K. Kato, A. Takefusa, H. Nakada, and M. Oguchi, Construction Scheme of a Scalable Distributed Stream Processing Infrastructure Using Ray and Apache Kafka, in Proceedings of the 34th International Conference on Computers and Their Applications (CATA 2019), pp. 368-377 (2019).

[14] Apache Hadoop, available at https://hadoop.apache.org/ (accessed June 1, 2019).

[15] Apache Storm, available at http://storm.apache.org (accessed June 1, 2019).

[16] Apache Flink, available at http://flink.apache.org (accessed June 1, 2019).

[17] Apache Spark Streaming, available at http://spark.apache.org/streaming/ (accessed June 1, 2019).

[18] M. A. Lopez, A. G. P. Lobato, and O. C. M. B. Duarte, A Performance Comparison of Open-Source Stream Processing Platforms, in Proceedings of 2016 IEEE Global Communications Conference (GLOBECOM 2016), 6 pages (2016).

Time-series Analysis using Idea-generation Support Groupware

Hideki Goromaru*, Jun Munemori**, Junko Itou**, and Takaya Yuizono***

*Secure Platform Group, Nippon Telegraph and Telephone Corporation, Japan
** Faculty of System Engineering, Wakayama University, Japan
***Human Life Design Field, Japan Advanced Institute of Science and Technology, Japan
goromaru@fw.ipsj.or.jp

Abstract - Risk-management techniques for identifying and analyzing risks regarding safety and security have recently been developed. However, the addition of security extends risk-management time by repeating safety, security and their harmonization. Brainstorming is used as a risk management method, some of which has already been electronized as idea generation groupware. Therefore, it is conceivable to apply idea-generation groupware for shortening multiple sessions in risk-management time. Before applying groupware for reducing risk-management time, it is necessary to investigate the characteristics of idea generation on groupware and determine how to apply it. In this study, we investigated the characteristics of the idea-generation process with the idea-generation support groupware GUNGEN. The results indicated that (1) there was no significant difference in the number of ideas generated per session between sessions of 1 hour or less and those longer than 1 hour (an average of 45 ideas were generated [standard deviation: 15], and about 30 to 60 ideas were generated in 80% of the sessions), and that (2) ideas were generated every 60 to 120 seconds on average, and about 90% of ideas were generated within 300 seconds.

Keywords: Risk management, safety, security, groupware, idea generation.

1 INTRODUCTION

Internet-of-Things (IoT) [1]and cyber physical system (CPS) [2] technology has recently spread to the industry which have mainly control systems such as automatic drive vehicles, thermal power plants. Society 5.0 [3], which is A human-centered society that balances economic advancement with the resolution of social problems by a system that highly integrates cyberspace and physical space[4]. Older control systems, which were specially-developed and were not standard, had used networks, communication protocols and equipment developed independently in the industry. The current systems have been transformed into standardized and generalized systems due to the spread of information systems. As a result of the links between devices or services becoming easier, new services can be easily implemented via the cloud or smartphones, reducing system-maintenance costs and improving convenience.

However, incidents and accidents, such as equipment failure and malware attacks, are likely to spread throughout a system and network. Therefore, there is a growing possibility that their spread will have national and global impact, for example, governments, banks, important institutions such as power plants, and infrastructure services will be threatened or shut down. Stuxnet [5]proved that even closed networks may be vulnerable to cyber attacks, and control systems that do not connect to a network are no longer safe. A control system is forced to take into account not only safety but also security; thus, various risk-management techniques regarding safety and security have been developed [5][6][7][8].

Risk-management techniques include many methods conventionally used for safety such as FMEA, FTA, and HAZOP, and those such as brainstorming [9] and Delphi are used for idea generation [8]. STAMP/STPA and FRAM have recently been proposed to cope with functional-resonance accidents, which are caused by abnormal signal generation due to noise mixing and resonance of multiple signals among complex system components [10]. As control systems are targeted by cyber-attacks, the number of methods regarding both safety and security have increased [5][6] , e.g., Security Informed Safety (SIS) [5][11] , which is the concept of handling safety based on security information.

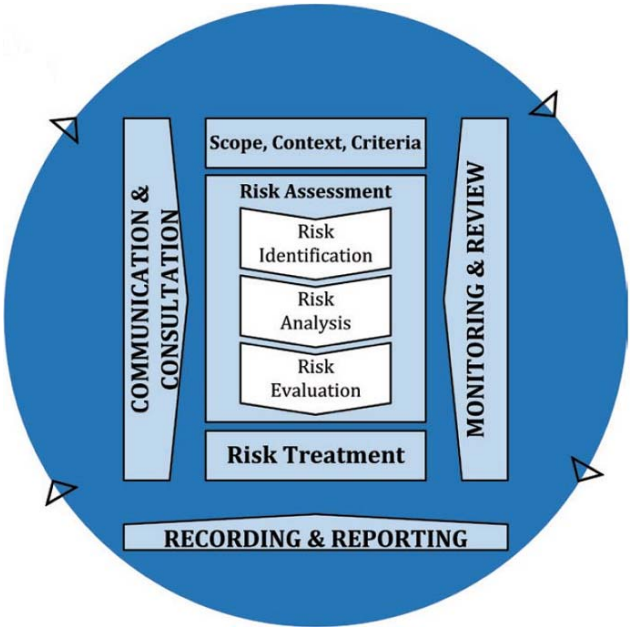


Figure 1: Process of Risk Management in ISO31000:2018

The process of risk management is shown in Figure 1 [12]. The steps in this process are as follows;

- Risk identification (detecting risks),

- Risk analysis (investigating the characteristics of risks),
- Risk evaluation (decision regarding the corresponding risks) and
- Risk treatment (thinking about countermeasures).

Most conventional risk-management techniques that include safety and security methods as well as safety-only methods include the risk-identification step. However, there are few risk-management techniques that include risk-treatment step [6][8]. Risk treatment requires creativity for formulating solution options (ideas) to address risk [12]. Therefore, to strengthen risk treatment, it is conceivable to use idea-generation support groupware that incorporates idea-generation methods for solving a problem. For example, GUNGEN, which is distributed idea-generation support groupware, is characterized by connecting personal computers (PCs) to networks and providing creative environments even remotely, such as in teleconferencing and video conferencing systems. This groupware makes it possible to solve the problem of remote locations and creativities, with conducting brainstorming at multiple remote locations via networks.

The risk-treatment time in a system development has to take into account not only the time for safety but also the time for security and the time for harmonization between security and security anew. Thus, with the addition of security and harmonization, the risk-treatment time has become longer. In general, various companies need to shorten system development time including risk-treatment time. However, the issue of reducing risk-treatment time has not been addressed in this groupware and the risk-treatment time using brainstorming in is not clear.

The length of many types of sessions including brainstorming is often about 1 hour, but less than 1 hour is preferred [13][14][15]. To address this issue, we need to know how ideas are generated in brainstorming sessions and reduce the idea-generation time. However, we do not know the idea-generation process. There are many studies that limit time in advance in brainstorming time series studies [16][17][18][19], but few studies have been verified until the idea is exhausted.

Therefore, we use the log-generation function of GUNGEN, which records the operations of session participants such as inputting idea with keyboard, moving the sticky notes freely with the mouse. GUNGEN records the time when an idea was generated and the name of the speaker and can clarify the idea-generation history, making it possible to determine who made what statement and when. Therefore, in order to shorten system developments time including risk-treatement time using brainstorming, we decided to follow the time sequence of how ideas are generated to find unnecessary steps in the idea-generaion process to reduce risk-treatement time using brainstorming (hereinafter referred to as “session time”).

We analyzed the generation of ideas over time and clarified the idea-generation process. We define terms in related research and their problems in Chapter 2, explain the analysis methods in Chapter 3, describe the experimental results in Chapter 4, provide discussion in Chapter 5, and conclude the paper in Chapter 6.

2 RELATED RESEARCH

The domain of system development for safety and security is called safety and security co-engineering [5]. The definitions of safety, security, and risk management in this domain differ from those in the current standard. This chapter gives the definitions of these terms for safety and security co-engineering and describes their relationships with idea-generation methods.

2.1 Safety and Security Definitions

In the current standard, safety is defined as "freedom from risk which is not tolerable” [20], and security (herein referred to as "information security") means "preservation of confidentiality, integrity and availability of information. Other terms, such as authenticity, accountability, non-repudiation, and reliability can also be involved" [21]. However, it is difficult to understand the difference between safety and security with these definitions. Abulamddi [10] describes the following two differences between safety and security.

(1) **Causal relationship**
- Safety is concerned with hazards arising from the system, potentially impacting the environment.
- Security is concerned with the risks originating from the environment, potentially impacting the system.

(2) **Intention or not**
- Safety is concerned with unintentional hazards.
- Security is concerned with intentional threats.

In this paper, we use the definition of difference (2).

2.2 Definition of risk management

Chockalingam et al. [6] gave similar definitions and included risk management as well as safety and security.

1. Safety is community as dealing with unintentional/non-malicious threats caused by natural disasters, technical failures, and human error.
2. Security is community as dealing with intentional/malicious threats caused by intentional human behavior.
3. Risk management plays a major role in dealing with both unintentional/non-malicious and intentional/malicious threats.

The definition of risk management in the current standard is "coordinated activities to direct and control an organization with regard to risk" [12] . This is different from the above definition of Chockalingam et al. We use the definition of risk management (3 above) including safety and security.

2.3 Relationship between risk management techniques and idea generation methods

Brainstorming is both a risk-management technique and an idea-generation method. It is also used in the idea-

generation phase of the KJ (Kawakita Jiro) method [22][1] . This section presents risk-management techniques and idea-generation methods.

(1) **Risk-management techniques and idea-generation methods**

Brainstorming is one of the risk management techniques [8], and be able to be used in Risk Identification and Risk Treatment.

(2) **Brainstorming and KJ method in idea-generation methods**

Guilford's Structure of Intellect theory includes "divergent-production" and "convergent-production" methods [23]. The two types of production methods are often used for classifying idea-generation methods. Divergent-production methods are used to generate multiple solutions to a problem, and convergent-production methods are used to deduce a single solution to a problem.

A well-known divergent-production method is brainstorming, and a well-known convergent-production method is the KJ method. The KJ method also includes brainstorming, which is a divergent-production method and is used in the divergent-production phase before convergent-production phase of the KJ method. GUNGEN supports idea generation in distributed environments by using the KJ method and enables brainstorming in the divergent-production phase of the KJ method.

Also, brainstorming is a type of session and an idea-generation method [15]. Idea-generation as almost same as session but it is difference that one participant can perform. In this paper, the following terms and definitions apply.

- Idea-generation method is a type of session, and one or more participants generate ideas in divergent-production phase and summarize ideas in convergent-production phase. In some cases, they repeat two phases or they only perform idea-generation.
- Brainstorming is an idea-generation method, and one or more participants mainly generate ideas in divergent-production phase.
- KJ method is an idea-generation method, and one or more participants perform brainstorming (KJ method mainly summarize ideas in convergent-production phase after brainstorming, but this article does not touch).

2.4 Related research on idea generation

This section shows related research on how ideas are generated in the divergent-production phase.

(1) **Idea generation without groupware**

Osborn [9] argued that ideas appear around every three minutes when brainstorming. According to a study by Munemori et al. [24], the idea-generation time is less than 1 hour on average, in the case of writing an idea on a sticky note and pasting the sticky note on a large white paper like a white board. In addition, it is not necessarily an idea generation, but generally, the length of a session is generally about 43 minutes for project sessions of around 5 participants, 78 minutes for those of around 8 participants, and about 1 hour for sessions of around 12 participants [15].

(2) **Improved idea generation without groupware**

There have been no reports on the time required for idea generation. In a study on brainstorming at the University of Chicago [9], the chairperson provided ideas and tips when it was determined that participants couldn't come up with any ideas and any ideas should be put forth. However, these are qualitative judgments by the chairperson without using quantitative methods such as time or number.

(3) **Idea generation with groupware**

According to the above study by Munemori et al. [24], sessions using idea-generation support groupware last less than 1 hour and 20 minutes on average. According to the research by Yuizono et al. [25], a large number of ideas are not always generated and monotonously decrease, or many ideas are not always generated at a certain time, when there are multiple participants.

(4) **Improved idea generation with groupware**

There have been no reports on the time required for idea generation. Donker et al. [26]proposed idea-generation support groupware that uses the 635 method (a type of brainstorming method also known as brain writing) [27], which 6 participants write down 3 ideas on a specific worksheet within 5 minutes and pass the worksheet onto the next member. With this method, it is possible to have the participants be anonymous, and to set the order of turning the worksheets with ideas in a fixed order or randomly. With computers, people can participate in creative group activities and conduct discussions without having to meet physically and without knowing each other.

Idea Expander [28] is idea-generation support groupware that uses photos when generating ideas. Participants brainstorm using the chat window of Idea Expander. This groupware supports idea generation of groups by displaying chat-related photos during brainstorming. GUNGEN-PHOTO [29]is idea-generation support groupware using the KJ method that further promotes idea generation using photos. This groupware has a shared work space for sharing discussions about photos to create ideas from the photos and a personal work space for entering ideas generated through discussions. It also enables the magnification photos and makes it easy to comment on the discussion and photos by displaying details during a project.

These studies on idea generation presented methods of reducing work load by automating certain processes or providing various methods for supporting idea generation. However, there was no mention of any unnecessary steps in idea generation for reducing session time. Therefore, it is necessary to examine the steps in the idea-generation process.

[1] KJ Method is a registered trademark of Kawakita Research Institute Co., Ltd.

3 ANALYTICAL METHOD

3.1 Experimental procedure

In this section, we identify the problems and issues involved in brainstorming in a near ideal environment with various biases removed.

(1) **Limiting survey target to number of ideas and time**

In brainstorming, one of the basic principles of idea generation is the thinking that quantity produces quality [9], that is the more ideas you produce, the better the quality of results. The number of ideas is most important when improving quality. We focused on the idea-generation process based on the number of ideas generated and time required to generate them and on confirming the presence or absence of unnecessary steps in this process (quality is for future study). Since the average session time including brainstorming is about 1 hour, as mentioned in Section 2.4, we compared whether there is a difference in the idea-generation process between 1-hour sessions and those longer than 1 hour.

(2) **Idea generation under normal and unusual circumstances**

Using the log-generation function of GUNGEN, we record the time when an idea was generated, the speaker of the idea, etc., to clarify the idea-generation process. Specifically, we investigated the following items based on the logs recorded with GUNGEN.

- Idea-generation process under normal circumstances (excluding the case such as inordinately long intermission and attends/leaves in the middle of the session)
- Idea-generation process under unusual circumstances (no ideas, few ideas, long sessions)

(3) **Tracking changes in idea generation over time**

The aim with this analysis was to understand the change from the start to the end of the idea-generation process (without adding any event such as giving hints and stimuli from outside). To determine how the idea-generation process changes over time due to the addition of new events is for future work.

(4) **Participants**

The experiments involved separate groups of 2nd to 3rd year students of Kagoshima University Faculty of Engineering. All participants in the same year knew each other. The goal was to acquire data from an environment as close to the ideal with less bias as possible. This is more difficult in corporate environments due to the following biases.

- People who have much experience in a particular job have much knowledge and can come up with many ideas. However, new employees with less experience cannot come up with many ideas because they have little knowledge.
- Many companies have a hierarchical structure. It is easy for the boss to give an idea because there are few people who would oppose him/her, and there are times when subordinates would hesitate to give their ideas.
- Session logs of actual companies are difficult to obtain because of the sensitivity of information.

In a student environment, it is possible for participants to hold sessions in a flat structure on a theme with similar knowledge. Therefore, the influence of the above biases can be reduced. Investigating actual corporate environments is for future work.

3.2 Experimental environment

(1) **Experimental system**

A group of three students sat in a row in front of three PCs with GUNGEN [24] installed (Figure 2). We focused on the following two functions.

- Brainstorming (divergent-production phase)
- Log generation (recording of participants' actions)

(2) **Theme**

Fixing a theme may make it difficult to come up with ideas depending on the degree of interest and knowledge of the participants, so themes were decided on topics that were easy for participants to discuss freely and give ideas. Therefore, the participants' interest and knowledge of the theme were as uniform as possible.

(3) **Session progress**

There was no presenter to maintain a flat relationship of participants. The session was for brainstorming, and students were free to speak and write. A staff member (a teacher in charge of the PCs or an undergraduate student) explained the experiment to the students, answered questions about PC operation, and went to the next phase according to the request from the participants.

Figure 2: Experimental environment

```
2943869394      KJOpen@!
2943870149      “showchat””Let’s start （Seguchi) ““@!”
2943870158      KJStart@!
2943870210      “showchat””Yes (Shiba) ““@!”
2943870232      “showchat””Understood (Shigeyama) ““@!”
2943870345      “showchat”” Do you aim for "ideal"? Do you aim for the "ultimate"?> all (Shigeyama) ““@!”
2943870512      ShowLabel”I can get an air ticket (Shigeyama)”
2943870619      ShowLabel”ATM/CD machine is placed”
```

Figure3: Example of log data

(4) **Experimental procedure**

The procedure of the experiment was as follows.

1. The staff member connected the PCs to the network just before the experiment and turned them on.
2. He/She explained the purpose of the experiment, brainstorming etiquette, e.g., do not criticize others' ideas and the quantity of ideas is more important than the quality of ideas, and how to use GUNGEN.
3. Students then decided the theme to be discussed.
4. After deciding on the theme, the students notified the staff member of the theme, sat down in front of three PCs (Fig. 2), and one of students entered the theme into GUNGEN.
5. Students pressed the start button on the screen to start the idea-generation process.
6. During this process, students input their ideas in the input field of the screen and pressed the idea button near the input field. A sticky note with the idea then appeared on the screen. Students could move the sticky notes freely with the mouse. In a similar way, students input their chats in the input field of the screen and pressed the chat button near the input field. The chat then appeared on the chat box.
7. The students generated ideas until ideas were exhausted.
8. The students notified the staff member when they all wanted to finish the idea-generation process.
9. The staff reconfirmed to the student the finishing of this process and concluded the brainstorming.

3.3 Analysis of logs

(1) **Theme**

Brainstorming on the 13 themes listed in Table 1 was caried out using GUNGEN, and the log data were analyzed.

For example, “Convenience store”, in the table represents the theme “Ideal convenience store”, e.g., it should be close to home and the cash register should be automated. "Food culture" represents the theme "the difference between food culture in foreign countries and Japan", e.g., the differences depend on the religion of a country. "Money" represents the theme "How does money affect people", e.g., whether debt destroys families, whether money is necessary in the age of digitization, and whether money can buy happiness.

(2) **GUNGEN log data**

Figure 3 shows an example of log data, which are explained in more detail below.

- The left column shows the time when an operation occurred. The unit time is seconds, and the cumulative seconds from January 1, 1902 are shown.
- The first part of the right column shows the operation type; "KJOpen @!" is system startup, "KJStart @!" is the start of the session start, “Showchat” is a generated chat, and "ShowLabel" is a generated idea.
- The second part of the right column shows the content. The letters indicate the data attached to the operation, and the parentheses indicate the operator's name.

The log data "ShowLabel" including the time and content of idea occurrence is important when analyzing.

Table 1: Experiment themes and number of ideas generated

Item No	Theme name	Number of ideas generated	Time (min)
1	Convenience store	43	80.6
2	New robot	23	71.0
3	Ideal future computer	70	106.0
4	Food culture	52	52.7
5	Ideal car	38	68.8
6	Ideal life	31	88.9
7	How to be the Ultimate Rich	37	31.7
8	About money	45	102.5
9	Ideal town	59	109.1
10	Ultimate fighting game	30	99.5
11	Ultimate city	79	85.7
12	Ideal city	40	107.4
13	Ideal university	40	52.4
Average		45.2	81.3

(3) **Analysis focus**

We focused on the time between idea generation, and the number of ideas generated within a certain time (seconds).

1. Distribution of total number of ideas in each session

Examine the distribution of the total number of ideas generated at each session. We examined this because we wanted to determine the frequency of the total number of generated ideas in normal sessions.

2. Intervals of generating ideas

Find trends in time between generation of ideas in seconds. This is to determine how often ideas are generated in sessions involving multiple participants.

3. Number of ideas per unit of time

Determine the number of ideas generated at a certain time interval, i.e., determine whether there is a difference between the generation of ideas when the session is short or long.

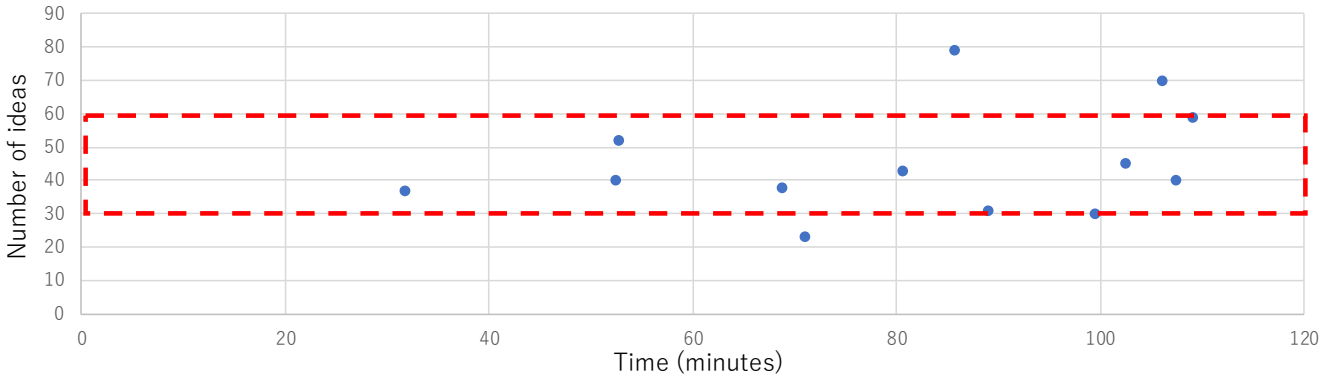


Figure 4: Distribution of session time and number of ideas

4 EXPERIMENTAL RESULTS

We first compared the results from sessions of less than 1 hour with those from sessions longer than 1 hour then investigated the number of ideas and the time interval of all ideas.

4.1 Number of ideas generated in sessions

The average number of ideas for the 13 sessions (themes) was 45.2, with a standard deviation of 15.4. Based on the mean and standard deviation, the total number of ideas was between 30 and 60 (45 ± 15) in 10 out of the 13 sessions, and almost 80% of sessions generated between 30 and 60 ideas as shown in Figure 4.

4.2 Idea-Generation Intervals

The average time between the idea generation for all the themes was about 110 seconds, and the median was about 70 seconds. Therefore, most ideas were generated between 60 and 120 seconds. Figure 5 shows 43% of ideas were generated within 60 seconds, 74% within 120 seconds, 83% within 180 seconds, 90% within 240 seconds, 93% within 300 seconds, and 95% within 360 seconds.

Figure 6 also shows the number of ideas generated for each time interval per theme;

19.2 (within 60 seconds),
13.8 (between 61 and 120 seconds),
4.1(between 121 and 180 seconds),
2.8(between 181 and 240 seconds),
1.3(between 241 and 300 seconds), and
0.8(between 301 and 360 seconds).

The number of ideas generated over 300 seconds was small.

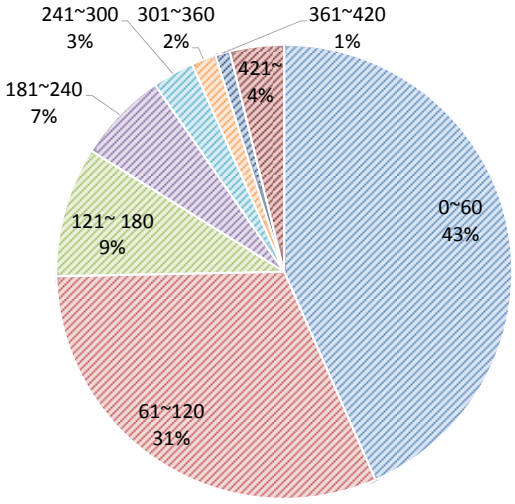


Figure 5 Distribution of intervals between generated ideas over time (seconds)

4.3 Comparison of sessions of 1 hour or less and those longer than 1 hour

Since most of the sessions took about one hour, as mentioned in Section 2.4 (1), we compared sessions of one

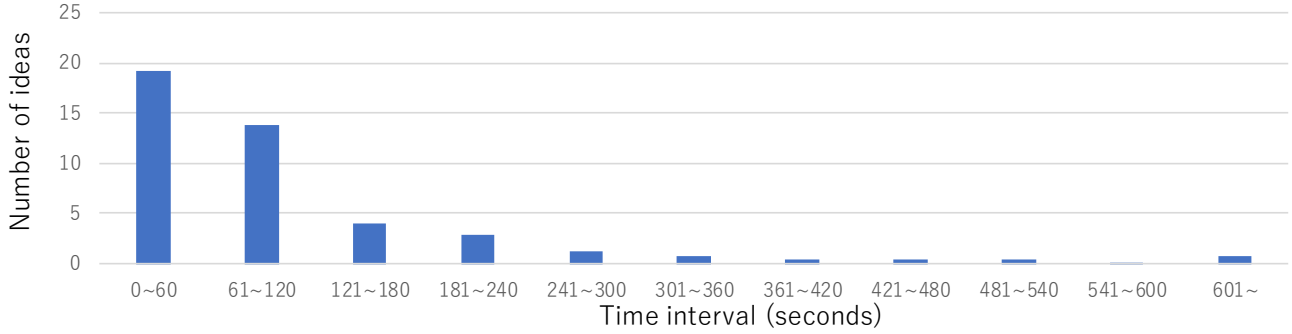


Figure 6: Number ideas generated for each time interval per session

hour or less with those over one hour. From the results presented in Section 4.2, we also examined the distribution in the number of average ideas generated in 300-second intervals.

(1) **Idea generation time (within 1 hour)**

Themes 4, 7, and 13 in Table 1 were targets. The average number of ideas generated for these three themes was 43.0 (Figure 7). Four or more ideas were generated every 300 seconds from the start of the session to 40 minutes into the session.

(2) **Idea-generation time (from 61 to 120 minutes)**

Theme 1, 2, 3, 5, 6, 8, 9, 10, 11, and 12 in Table 1 were the targets. The average number of ideas generated for these themes was 45.8 (Figure 8). Immediately after the start of the session, but this number per 300 seconds decreased over time.

5 CONSIDERATION

(1) **Number of ideas per idea-generation was between 30 and 60**

The number of ideas per idea-generation was 45.2 (average deviation 15.4) on average, and in 10 out of the 13 sessions, 30 to 60 ideas were generated (45 ± 15). Kawakita [30] stated that the number of ideas was about 30 to 60 per idea-generation in 1,000 or more practical cases. The average number of ideas per idea-generation, which is from the results of past student experiments by Yuizon et al. [31], was about 30 (1 participant), 40 (2 participants), and 50 (3 participants). This means that our results are consistent with those of these past studies.

(2) **Average number of generated ideas was between 60 to 120 seconds, and approximately 90% occurred within 300 seconds**

The average number of generated ideas was between 60 and 120 seconds, 84% was within 180 seconds and about 93% within 300 seconds. According to Osborn's report, most ideas are generated within 180 seconds in a corporate environment [9]. In our student environment, about 84% of ideas were generated within 180 seconds, which agrees with Osborn's results.

(3) **Number of ideas does not change even during long sessions**

The total number of ideas for one theme was 43.0 in a 1hour session and 45.8 in a session longer than 1 hour. The difference between the number of ideas generated for sessions over 1 hour and for those less than 1 hour was examined using Student's t-test, which is applicable even when the number of samples is small. There was no significant difference as $t(11) = 0.25$, $p > .05$. Therefore, there was no significant difference in the number of ideas generated even when the session time was extended.

The average number of ideas generated did not change, but there was a difference between the sessions less than 1 hour and those over 1 hour for ideas generated at 300-second intervals. A session less than 1 hour generated four or more stable ideas in 300-seconds intervals from the start of the session until 40 minutes. However, the number of ideas generated at 300-second intervals for sessions over 1 hour gradually decreased from the start to the end of the session.

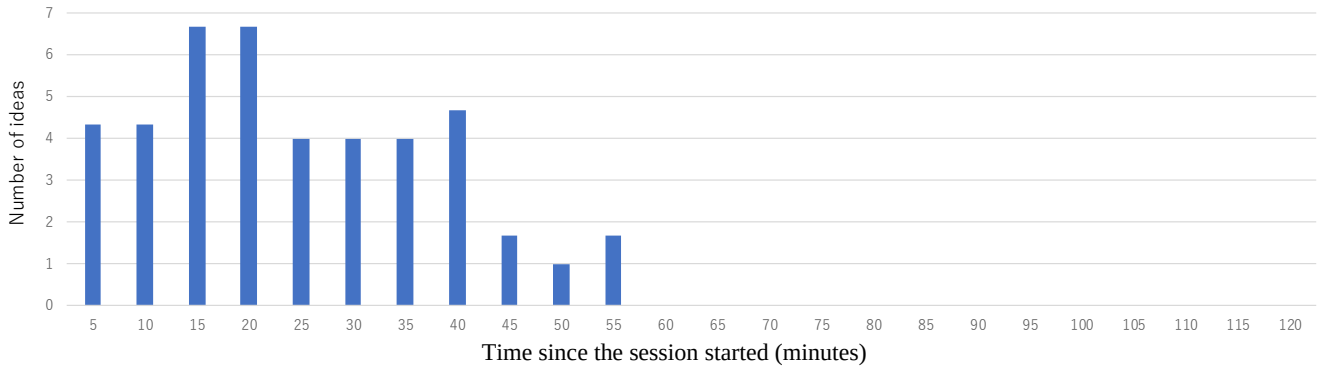


Figure 7: Distribution of average number of ideas generated in 5-minute units for sessions of 60 minutes or less

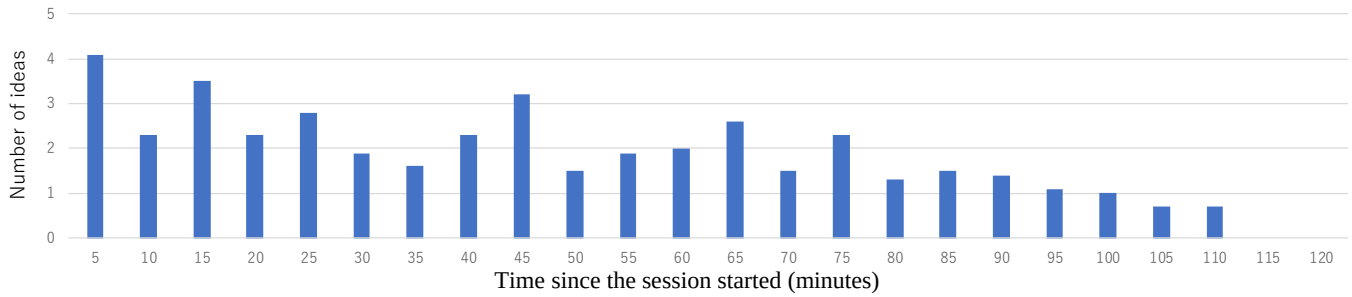


Figure 8: Distribution of average number of ideas in 5-minute units for sessions between 61 and 120 minutes

(4) **Application example based on the results**

As mentioned in (3), if the number of 300-second ideas is less than 4 and continuous, it is likely that the session time is more than one hour. As mentioned in (2), if the next idea is not generated within 300 seconds of the last idea being generated, it may not be generated during the session. As described in (1), it may be determined that it is better to continue the session if the total number of ideas is less than 30 and finish the session if the total number of ideas is more than 60. Based on these results, idea generation support groupware needs the following requirements in order to realize a session within one hour.

1. Function to count the number of ideas at 300-second intervals from the start of session (In addition, if the number of 300-second ideas is less than 4 and continuous, the function takes some action to promote idea generation.)
2. Function to count the time since the last idea was generated (In addition, if the next idea is not generated within 300 seconds of the last idea being generated, the function takes some action, e.g., prompting idea generation, warning to end or continue the session.)
3. Function to count the total number of ideas from the start of session (In addition, if the next idea is not generated within 300 seconds of the last idea being generated and the total number of ideas is more than 60, the function takes some action to promote to end the session.)

Figure 9 [32] shows an example of an idea generation support groupware system with functions 2 and 3 above to monitor the number of ideas and the time taken to generate them. If the number of ideas generated within 300 seconds of the last generated idea is 0, the system displays a hint for generating a new idea. If the number of ideas generated within 300 seconds of the last generated idea is 0 and the total number of ideas generated exceeds 60, then the system switches to the next phase (for example, the convergent production phase of KJ method).

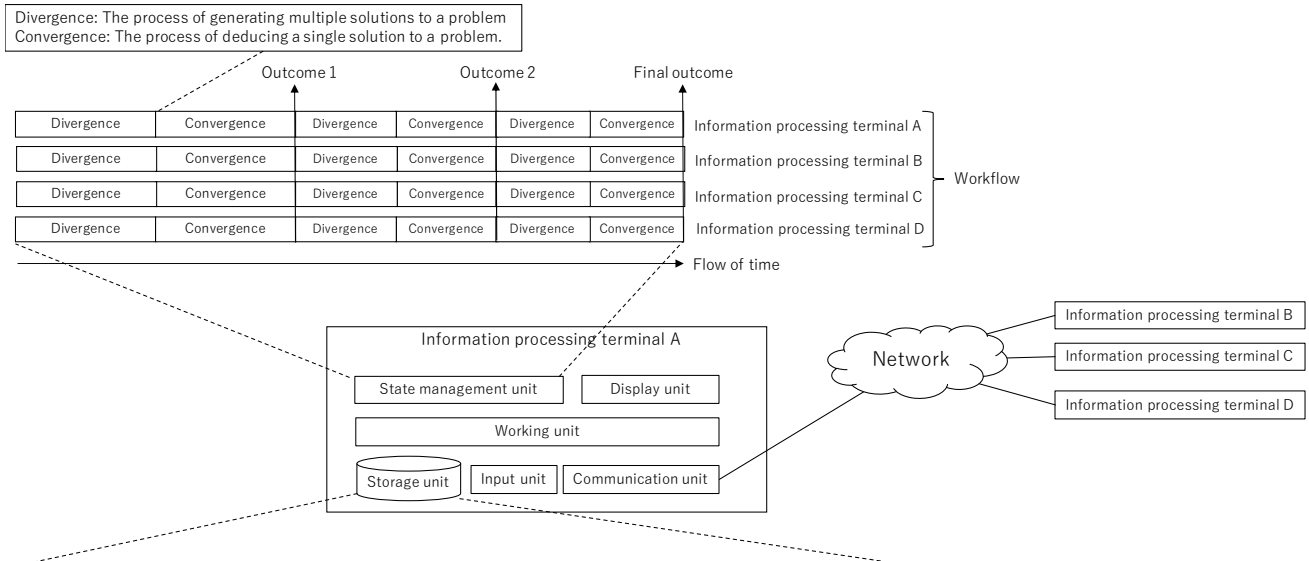
This approach is similar to the chairperson's operation in brainstorming at the University of Chicago [9]. However, the chairperson's operation is based on the chairperson's qualitative judgment without using quantitative methods such as time or number. The difference with this approach is whether to judge quantitatively.

6 CONCLUSION

Brainstorming sessions were conducted in groups of three student participants, and the idea-generation process was recorded using the log-generation function of GUNGEN. We found the following.

- The number of ideas per session was the same for sessions of 1 hour or less or sessions longer than 1 hour. The overall average number of ideas generated per session was about 45 (standard deviation is about 15), and almost 80% of sessions generated between 30 and 60 ideas.
- The average number of generated ideas occurred every 60 and 120 seconds, and about 90% of ideas were generated within 300 seconds.

When brainstorming using GUNGEN in our analysis, the number of ideas generated was almost 30 to 60 even when the session continued for a long time. Since the total number of ideas generated in a session did not change significantly over time, the goal is to guide the participants to finish the



Conditions	Process
If the next idea is not generated within 300 seconds of the last idea	Display hints (text or diagrams) to prompt ideas
When the number of ideas exceeds 60	Display warning to advance to the next phase

Figure 9: Application example for idea generation support groupware which has divergent-production phase and convergent-production phase (e. g., KJ method) based on the experiment results

session within 1 hour to promote efficient idea generation. Thus, for more than 300 seconds after idea generation, when it is determined that few or no more ideas would be generated, a prompt needs to be given to encourage action or promote further idea generation.

In the future, we plan to examine and verify the following points.

- The effect of shortening system development time including risk-treatment time (session time) by the new approach.
- More effective ways to encourage the creation of ideas (for example, use of images).
- More effective ways to accelerate the transition to the next stage.

REFERENCES

[1] Brian Russell et. al., Practical Internet of Things Security, Packt Publishing, 2016.

[2] Khaitan et. al., Design Techniques and Applications of Cyber Physical Systems: A Survey, IEEE Systems Journal, 2014.

[3] Society 5.0, https://www.gov-online.go.jp/cam/s5/eng/index.html, last accessed 2019-01-18.

[4] Panel2: Society 5.0 in Depth, https://www.jcci.or.jp/Panel2%20Jetro%20Mr%20.Mitsumata.pdf, last accessed, 2019-01-18.

[5] MERgE ITEA2 Project. Recommendations for Security and Safety Co-engineering, MERgE ITEA2 Project # 11011, 2016.

[6] Sabarathinam C. et. al., Integrated Safety and Security Risk Assessment Methods: A Survey of Key Characteristics and Applications, Cornell University Library, 2017.

[7] Yulia C. et. al., A review of cyber security risk assessment methods for SCADA systems, Elsevier, computers & security56 (2016) pp.1–27, 2016.

[8] Risk management – Risk assessment technique, IEC/ISO 31010: 2009.

[9] Alex F. Osborn, Applied Imagination; Principles and Procedures of Creative Thinking, Creative Education Inc., 1963.

[10] Mohammed F. et. al., A survey on cinques requirements for integrating safety and security engineering for cyber-physical systems, JCSES, Vol.7, No.6, 2016.

[11] Elena Lisova et. al., Safety and Security Co-Analyses: A Systematic Literature Review, IEEE SYSTEMS JOURNAL, 2018.

[12] Risk management – Guidelines, ISO 31000: 2018.

[13] Survey on "a moderate time" at the beginning and end of the Heisei period (Japanese only) https://www.citizen.co.jp/research/time/20181115/01.html, last accessed 2019-01-18.

[14] NTT Com Research: Opinion Survey on Meetings (2009) (Japanese only), http://research.nttcoms.com/database/data/001053/, last accessed 2019-01-18.

[15] Karin Niemantsverdriet et. al., Recurring Meetings. An Experiential Account of Repeating Meetings in a Large Organization, December 2017 Proceedings of the ACM on Human-Computer Interaction: Volume 1 Issue CSCW, 2017.

[16] Taylor, D. W. et. al., Does group participation when using brainstorming facilitate or inhibit creative thinking?, Administrative Science Quarterly, 3, 23-47, 1958.

[17] Diehl, M., Stroebe, W., Productivity Loss in Brainstorming Groups, Personality and Social Psychology, 53(3), 497-509, 1987.

[18] Paulus, P. B., & Dzindolet, Social influence processes in group brainstorming, Personality and Social Psychology, 64, 575-86,1993.

[19] Joao Pissarra et. al., Brainstorming Reconsidered in Computer-Mediated Communication and Group Support System Context, ACM, ISDOC'12, 2012.

[20] Safety Aspects-Guidelines for their inclusion in standards, ISO/IEC Guide 51: 2014.

[21] Information Technology-Security Techniques-Information security management systems-Overview and vocabulary, ISO/IEC 27000: 2014.

[22] Affinity Diagrams, https://www.ieee.li/pdf/essay/affinity_diagrams.pdf, accessed 2019-01-18.

[23] Guilford, J. P. The Nature of Human Intelligence, McGraw-Hill, 1967.

[24] Jun Munemori et. al., GUNGEN: groupware for a new idea generation support system, Information and Software Technology 38 (1996) pp.213-220, 1996.

[25] Takaya Yuizono et. al., A Study of Relation between Cooperative process and Result, on the Distributed and Cooperative KJ Method Using Groupware for a New Idea Generation Support System. Journal of IPSJ, Vol. 39 No.2, pp.424-437, 1998.

[26] Donker, H., Computer Support of Creativity Processes in Teams, ISBN: 978-972-8924-39-3 (2007)

[27] Donker, H., Supporting Creativity in Virtual Teams, CollabTech 2008, pp.182-8, 2008.

[28] Wang, H. et. al., Idea Expander: Supporting Group Brainstorming with Conversationally Triggered Visual Thinking Stimuli, Proc. The 2010 ACM Conference on Computer Supported Cooperative Work, CSCW2010, pp. 103-106, 2010.

[29] Kokogawa et. al., The Effect of Using Photographs in Idea Generation Support System, Journal of Information Processing, Vol. 21, No. 3, 2013.

[30] Jiro Kawakita, KJ method - to speak in chaos (Japanese only), Chuoukouron-sha, 1986.

[31] Takaya Yuizono et. al., Investigation of the Cooperative KJ method for Collective Intelligence with Groupware KUSANAGI for a New Idea Generation. IPSJ-JNL, Vol.53 No.11, pp.2635-2648, 2012.

[32] Hideki Goromaru et. al., Identification Support Device, Identification Support Server, and Identification Support Method (Japanese only), Patent No. 5645026, 2014.

Session 3:
Network and Security
(Chair: Kozo Okano)

Victims Information Management System for Large Scale Disaster

Kosuk Akasaka [*], Takayo Namba [**], Masao Isshiki[*] and Keiichi Abe [*]

[*] Kanagawa Institute of Technology, Atsugi City, Kanagawa, Japan
[**] Tokyo Metropolitan University, Hachioji City, Tokyo, Japan
abe@he.kanagawa-it.ac.jp

Abstract –At the time of past large-scale disasters such as the Tohoku Earthquake in 2011, it took a considerable amount of time to collect information on the victims.

This paper propose a system that installs Our Refuge Management System in refuges all over the country, and uses cloud servers to centrally manage the disaster situation and relief needs in each refuge area. We call this system the "Victims Information Management System (VIMS)" and are developing it. In this paper, we describe the contents of the research and development project.

Keywords: Disaster, Refuge Management System, Victims, Persons requiring special care, ICT, Solar Panel

1 INTRODUCTION

At the time of past large-scale disasters such as the Hanshin-Awaji Earthquake in 1995, and the Tohoku Earthquake in 2011, it took a considerable amount of time to collect information on the victims [1]. In particular, refuge administrators couldn't capture relief needs such as the presence of refugees who need special care and their allergy information. Also, due to the disruption of the infrastructure at the time of disaster occurrence, managers were unable to make the most of the Information and Communication Technology (ICT) and its power, and they could not collect and transmit information on the victims promptly. [2] [3]

In addition, there were many victims who could not cope with the refuge environment and suffered from health problems. [4] [5]

Therefore, we developed the Refuge Management System (RMS) that consistently manages the health condition of victims by collecting information on disaster victims and creating an evacuation list using ICT even at the time of the disruption of lifelines such as communication and electricity. [6] [7] The final goal of this study is to develop a system that installs RMS in refuges all over the country, and uses cloud servers to centrally manage the disaster situation and relief needs in each refuge area. We call this system the "Victims Information Management System (VIMS)" and are developing it. In this paper, we describe the contents of the research and development project and the realization of the system to manage personal information and health condition of the victims in the refuge currently under development.

2 RELATED RESEARCH

After the Tohoku Earthquake, research on victims' support systems using ICT is underway by many researchers, companies and universities in various places. [8]

Similar representative existing technologies include "Earthquake disaster recovery support system" provided by Microsoft [9] and "information center victim support system" provided for free by Nishinomiya city [10]. These systems collect information such as the number of victims, the proportion of men and women, safety information, etc. On the other hand, these systems are not designed to pick up and manage the needs information of persons requiring special care (patients with allergic diseases or intractable diseases, disabled people, elderly people who need nursing care, pregnant women, etc.) within the refuge. In addition, these systems have the problem that they cannot detect and manage victims who have become ill due to a long-term refuge life.

Therefore, the problems of the various disaster victim supporting systems are as follows.

① The existing disaster victim support systems cannot operate when there is a power and communication infrastructure disruption immediately after the disaster occurs.
② The existing systems cannot adequately collect the needs of persons requiring special care.
③ It is difficult to manage the presence of victims among multiple refuges.
④ The QoL of the victims declines in refuge life.
⑤ It is impossible to centrally monitor damage situation of each refuge and relief needs in real time immediately after the disaster.

The VIMS proposed in this paper aims to solve problems ① to ③.

3 VICTIMS INFORMATION MANAGEMENT SYSTEM (VIMS)

3.1 Outline of VIMS

In refuge support, the measures to be taken in refuges change with the passage of time from the occurrence of the disaster. For example, even if the power and communication infrastructure are cut off immediately, it is necessary to quickly grasp the situation of the victims in the refuges and inform external organizations of the situation after the occurrence of a disaster. On the other hand, there are many victims who are forced to live in refuges even after the restoration of infrastructure is completed and a long time has passed since the disaster occurred. For such victims, refuge managers need to manage their health status and prevent infections, economy class syndrome and so on. Furthermore, it is also necessary to grasp the damage situation regarding the collapse of the houses owned by the victims with the passage of time.

Therefore, in this research, as shown in Fig. 1, research and development are being carried out in the following four phases aiming for the realization of the final goal, VIMS. We assume that Phase 1 of VIMS will collect victim information, send it to Disaster Countermeasures Headquarters, and manage it in shelters immediately after a disaster occurs. We suppose in Phase 2, there are determined the transported destination and refuge spaces of the victims appropriately from the disaster situation and medical condition of them. We assume that in Phase 3 we will manage the health status of victims in long-term refuge life in real time. In Phase 4, we will devise a system that is able to monitor centrally the disaster situation and relief needs in a large area refuge. The main feature of VIMS is the support and management of victims in single system, from the immediately after a disaster occurs to the long-term shelter life. This feature is not found in the existing technology, so it can be said to be a great strength of this research.

Specifically, we will develop the Refuge Management System (RMS) corresponding to Phase 1 to Phase 2 shown in Fig. 2 and aim to install the proposed RMS at evacuation sites all over the country. The final goal is to realize a system that visualizes the disaster situation of refuges nationwide and the relief needs information of the victims by listing the victim information collected by the RMS as a cloud server by wireless communication etc. and mapping it with map information. In this paper, we describe the research results of Phase 1 and Phase 2 of this research project and the overall concept of the information management system of the victims in the refuge.

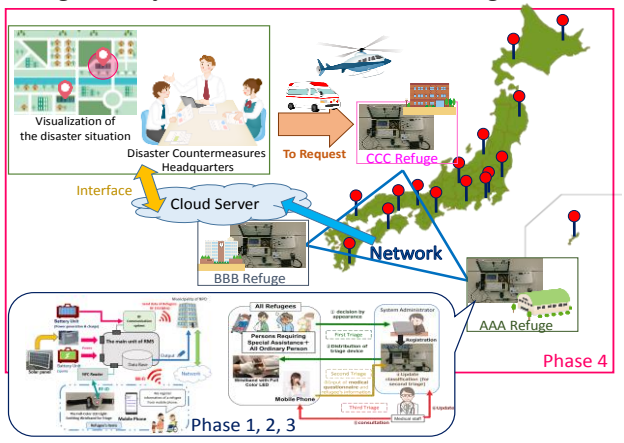


Fig. 1 Outline diagram of VIMS

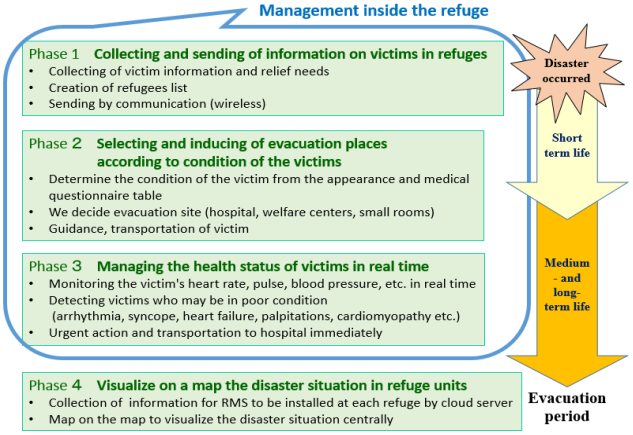


Fig. 2 Phased timeline

4 REFUGE MANAGEMENT SYSTEM (RMS) DEVELOPMENT CORRESPONDING TO PHASE 1

Chapter 4 describes the design concept and system development for the RMS corresponding to Phase 1 of the VIMS research and development.

4.1 Concept of RMS for phase 1

We designed RMS with the following concept.

I. The system user

The administrator of the RMS is preferably a civil servant etc. Because, the administrator deals with victim information.

II. Collection and management of victim information

We assume that the installation site of the RMS has a capacity of about 1000 individuals per refuge, such as schools and community centers.

III. Power supply of RMS

Lead battery units charged with electric current generated by solar panels are used as the power source of the RMS. this makes the RMS available even when the infrastructure is not working.

IV. Data collection content

The contents of data to be collected shall be the name, address, gender, age, nationality, emergency contact information, damage situation, and stockpile status of evacuation supplies, items requiring long-term care (pregnancy, chronic illness, and allergy). These items of information concerning the necessary nursing care were added to the questionnaire because they were actually used at the refuges of the Tohoku Earthquake. Additional items to be entered include a User Identifier (UID) to identify the victim and a questionnaire to confirm the health condition. These information is input by victims themselves or staff members who support refugees.

V. Identification for victims

As a method to identify each victim, we use the UID of NFC (Near Field Communication) or RF-ID information with a full color LED wristband distributed in the refuge. By using these data to manage entry and exit in the refuge, we can grasp the situation of the victim's presence in the refuge and track transfer of other victims to other refuges.

VI. How to send victim information

We consider two patterns of transmission methods for victim information when the communication infrastructure is not working. The first method is to connect a wireless unit (simple type radio, amateur radio) to the RMS. By this method, we can send information to the national disaster response headquarters using character data in JSON format or CSV format through any communication protocol. Another way is to save it in a USB flash memory and take it out. After the infrastructure is restored, the system can acquire the victim information through the Internet.

VII. Communication method between RMS and various information terminals

When inputting victim information, Wi-Fi is used for the communication between the information terminal such as a smartphone and the RMS. If the number of connections with information terminals is insufficient, we can increase the number of repeater devices.

4.2 Refuge Management System

In R & D for Phase 1, we prepared an outline design of the refuge management system. The outline is shown in Fig. 3. A commercially available microcomputer board Raspberry Pi Model B (RPI) was used for RMS hardware. We installed an embedded Linux OS (Raspbian) on this hardware and designed a system that can launch Graphical User Interface (GUI) applications. The reason for adopting embedded Linux on hardware is to save power of the RMS itself. The power consumption of an ordinary notebook PC is about 40 ~ 50 [W], which can be reduced to about 1/10 by using RPI. In addition, with regard to the information input of evacuees described in Design Concept Ⅳ, it is necessary to allow for the input of refugee's information from a web browser on a smartphone or mobile phone owned by the victims. Therefore, we constructed a web server using apache2 and developed a PHP application for a disaster victim information input screen. As shown in Fig. 3, the refugee's information registration application was developed using a server-side application that does not depend on the client environment so that victim's information can be input from the WWW browser installed on the client side.

Victims are requested to input personal information such as the name, address, gender, and age using the input screen shown in Fig. 4. In addition, we have added selective buttons asking whether or not the respondent is a pregnant woman, a person with a disability, or requires nursing care, etc., as well as check boxes to input food allergy determined by law. (Wheat, egg, milk, buckwheat, shrimp, crab, peanut, etc.) [14] In consideration of a more detailed description and answer which doesn't correspond to any items, we have also provided a text box for free description. By summarizing such information, it is possible to create a victims roster including the relief needs within the RMS. Based on the above, we will realize a system that picks up the relief needs of the affected people, including those with special needs, and will eventually solve Problem ② in Chapter 2.

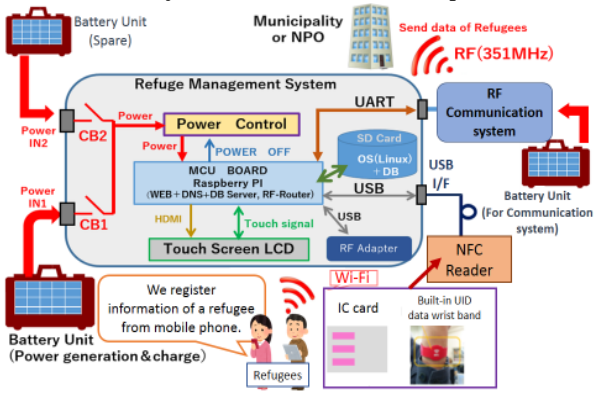


Fig. 3 RMS corresponding to Phase 1

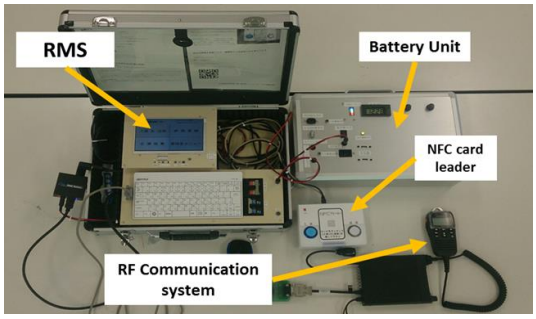


Fig. 4 RMS prototype

By the way, the system of Phase 1 assumes operation at an actual refuge. Therefore, the system must be able to operate even when the power infrastructure is not working. Therefore, we will develop a battery unit with a solar panel and a lead storage battery as the power supply of the RMS, and use two or more such battery units to operate the system. The operation method is described in the next section.

In addition, we developed a method for managing the evacuation using the RMS in preparation for the case where victims irregularly enter or exit from the refuge. We conceived the idea of using NFC cards which the victims usually carry with them. Entering and exiting of refuges for victims is done by holding the NFC card over the NFC card reader. By associating the UID of NFC with the already registered victim information in the RMS, it is possible to quickly manage the entry and exit of each victim.

The refugee's roster created by the RMS is organized as an electronic file, and can be saved in a CSV format and a JSON format on an electronic medium such as a USB flash memory. As a result, the data list of evacuees can be taken out to external facilities such as the disaster response headquarters. In addition, amateur radios and simplified radios which are available for the communication even during a large-scale disaster can be connected to the RMS with a serial cable to send a relief request to the disaster response headquarters.

4.3 RMS PROTOTYPE TO PHASE 1

Fig. 4 shows an outline of the proposed RMS prototype system. The RMS consists of an RMS server, a battery unit, an NFC reader panel, and a wireless communication device. The main power of the RMS server is supplied from the battery unit. By connecting this unit to a solar panel, the generated electricity can be stored in the lead battery. The lead battery unit is made of an attachment case containing a lead storage battery and a thin solar panel, making it easy to carry. The battery unit requires a minimum charging time of 5 hours under fine weather condition. In order to operate the RMS power supply in the long term, we developed two lead battery units which should be used alternately. By alternately performing these operations, we enabled the RMS to operate for a long time. We actually operated the system using this method for a week. As a result, there was no problem with the RMS MCU. Also, we confirmed that the operating time can be extended without shutting down the system power.

This will allow the RMS to operate even when the power and communication infrastructure are disrupted, and we believe that we can solve Problem ① in Chapter 2.

5 RMS DEVELOPMENT CORRESPONDING TO PHASE 2

Chapter 5 describes the contents of the RMS development corresponding to Phase 2 of the VIMS research and development. Specifically, we developed a system to perform color coding according to the physical and mental condition of the affected people, including persons requiring special care. In addition, we will explain the method to guide victims to appropriate living spaces in the refuge based on judgment from their physical conditions.

5.1 Triage for persons requiring special care

As prior research corresponding to phase 2 of this paper, M. Ohara et al. proposed "Triage for persons requiring special care"[11]. Table 1 shows the criteria of triage for persons requiring special care proposed by Ms. Ohara. The table provides decision criteria for the leader of the residents to decide the room allotment for persons requiring special care when entering the evacuation refuge and the priority of transfer to the welfare evacuation refuge. It is different from the Simple Triage and Rapid Treatment method. The proposed triage method uses four colors: orange, pink, purple and blue. Because of this, the colors do not mix, so the two triage methods can be used simultaneously.

5.2 Problem of the proposed triage and solution

This time, we considered an RMS that incorporated the "triage method for persons requiring special care " proposed by Ms. Ohara. In small-scale refuges, medical staff may not be present. Even under such circumstances, we consider it is necessary to use the proposed triage to make classification decisions according to the condition of the affected person smoothly. So, we have devised a system that will determine color classification automatically to some extent even without medical staff, by having the victims respond directly to an electronic questionnaire. Furthermore, it is necessary to consider the privacy protection by the burden on mental and psychological stress by coloring the victims. Therefore, we designed a wristband for PRSC triage with a full color light emitting LED so that the division color of the triage could be known. By having this for all the victims, it is possible to identify the target person by lighting the LED only when necessary. As similar electronic triage devices is being conducted currently at various companies and universities in emergency medical care [12] [13]. On the other hand, the “wristband devices for persons requiring special care triage” we propose is intended for the need-minded person at a shelter where medium to long time has passed since the occurrence of the disaster. Therefore, the application is different from conventional electronic triage equipment.

Table 1 Criteria and classification of triage for persons requiring special care [11]

color classification	Class	Decision Criteria	Area
1 (Orange)	Person who needs urgent treatment	Injuries with bleeding Fever, Diarrhea, Nausea, Vomiting, Dialysis patient , Oxygen suction patient	Hospital
2 (Pink)	Person who needs full assistance in daily life	Person who can't eat, walk or excrete independently Bedridden person	Welfare center
3 (Purple)	Person who needs some assistance in daily life	Pregnant women Infants under 3 years of age Half paralyzed person People with intellectual disabilities Wheelchair user	Small room (classroom etc.)
4 (Blue)	People who can act without problems	Person who can walk independently	Large room

In particular, our wristband devices has a power switch and victims can turn on the LED for it at any timing. In other words, when not using the system, turning off the system protects the evacuees' privacy. In addition, by using this wristband type device, it is possible to instantly identify the victims by the emission color of the LED even in a small space. Therefore, refuge staff can quickly guide the evacuees to the appropriate evacuation space. In addition, it is necessary to consider victims who become sick during the life in the refuge after a large-scale disaster. Therefore, we thought that a system that could check the health status of victims in real time was necessary.

Therefore, in the phase-2 system of this proposal, we have devised a system of additional functions combining the "triage of users requiring consideration" proposed by Mr. Ohara with the RMS of phase 1 and the full color LED light emission wristband described above.

5.3 Operation of VMT

The method proposed by Ms. Ohara is a two-step classification method including the "primary triage" to judge by appearance, followed by the "secondary triage" based on the decision of medical staff [14]. On the other hand, the triage we propose is a three-stage system, which we call the Victims Management Triage (VMT). Fig. 5 shows the VMT system. Section 5.3 describes how to use the VMT.

Fig. 5 shows the management method in the refuge using the VMT. This system should be used at least three times to perform three types of triages which we call “First Triage”, “Second Triage”, and “Third Triage”. "First Triage" is an initial classification by refuge administrators and staff. "Second triage" is done by the evacuees themselves by entering the questionnaire. After that, the method of making a judgment based on medical examination by a doctor etc. is called "Third Triage". We describe the method of each triage below.In the first triage, the refuge manager and staff will determine the triage classification from the appearance for all victims entering. At this time, the staff will distribute

wristband devices for VMT to all victims. The reason for judging the status of evacuees only by appearance in the first triage, is to prevent congestion at the entrance and reception of the refuge.

After several hours from the start-up and looking at the people inside the refuge, the second triage is performed at an arbitrary timing. In the second triage, victims can enter mental and physical information by selecting the item of the e-questionnaire table from the RMS registration application shown in Fig. 6. Concretely, they register their personal information (name, address, disaster situation etc.) in the RMS, and at the same time, make an inquiry by self-report which inputs the health condition such as the presence or absence of injury or disease.

The third triage is to re-determine the classification from the information obtained by the second triage and from the consultation of victims by a patrol doctor and medical staff. Because the conditions of all victims change with the long-term life in the refuge, the e-questionnaire table can always be changed by the victims themselves, and the color of the triage can be instantly updated according to the condition of the person requiring consideration or the judgment of the traveling doctor.

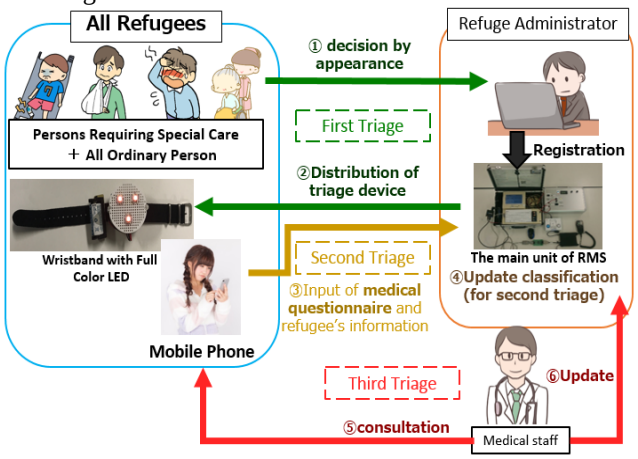


Fig. 5 Victims Management Triage

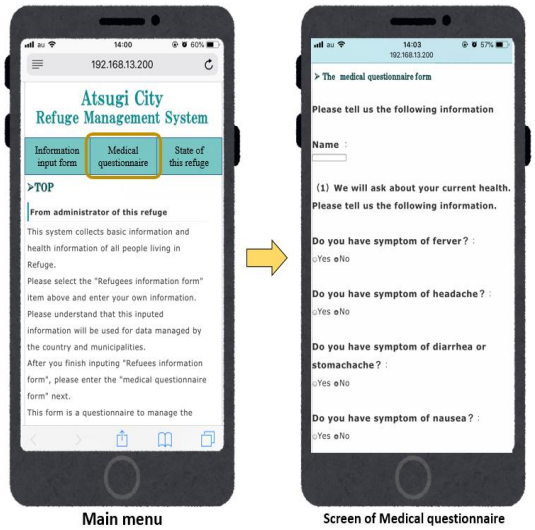


Fig. 6 Screen of Medical questionnaire

At this time, the refuge administer can easily find the person who corresponded to the 4 triage colors by turning on the switch of their wristbands. As a result, the staff will be able to quickly guide the victims to the appropriate evacuation space. On the other hand, when the transport destination such as a hospital or a welfare facility is not ready, the victim turns off the light of the wristband devices for VMT and stands by at a waiting place or the like until the preparation is completed.

5.4 e-questionnaire table from the RMS

Fig. 86 shows the registration screen of the electronic questionnaire. By answering this electronic questionnaire, the RMS automatically performs the VMT. Table 2 shows triage classification criteria for the answers of the electronic questionnaire.

As a specific criterion for the triage from e-medical questionnaire, the e-medical questionnaire form has four question items: "MQ 1: health information", "MQ 2: presence / absence of assistance in daily behavior", "MQ 3: prepartum / postpartum women", and "MQ 4: existing medical treatment (such as a chronic illness) ". We will explain each check item in detail. In MQ 1, the respondent is asked to select any symptom that applies to him/her from the seven items of fever, headache, abdominal pain (diarrhea), nausea, dizziness, cough (sputum), and runny nose. MQ 2 is a question about whether the respondent requires assistance in any of the three items of walking, meals, and excretion. The answer is selected from three options: "no problem", "some actions require assistances", and "inoperable without assistance". MQ 3 is an item that asks whether or not the respondent is pregnant or has any infant, and if applicable, whether the pregnancy period has exceeded nine months. MQ 4 is an item that asks whether a medical treatment such as dialysis and medication is necessary, and whether the treatment is necessary within 3 days. Also, in consideration of detailed descriptions and answers that do not correspond to any items, a text box for free description is provided. The RMS automatically determines the color of the division into the triage based on the criteria of Table 2 from the answer result of this e-examination table. It is thought that this will solve Problem ③ " The QoL of the victims declines in refuge life " mentioned in Chapter 2.

5.5 Prototype system of VMT

An outline of the VMT prototype proposed in this paper is shown in Fig. 7. The VMT is a system that adds the proposed VMT operation functions and the LED wristband to the RMS of Phase 1. In order to identify the color-coded persons, the manager distributes a full-color LED light-emitting wristband to all evacuees. Evacuees will be able to guide the evacuation site smoothly by emitting LEDs as needed.

The LED wristband used an AT mega328P for the wristband Micro Control Unit (MCU) and was developed in Arduino-like C language. The IEEE 802.15.4 standard was used for communication with the RMS main unit. The

reason for using this standard is that it can connect with many nodes (theoretically about 40,000 or more) compared to the Wi-Fi standard. The RMS acts as a coordinator, and the wristband acts as a device.

Fig. 8 shows the communication procedure between the RMS and the LED wristband. We designed the RMS to transmit a request command to the LED wristband so that the emission color of the LED can be changed as needed. We defined the command length of this system as 10-byte character data. The 10-byte character data includes 7 bytes as the device number, 1 byte as the color code, and the remaining 2 bytes as the delimiter (combination of CR and LF). The device number is the unique ID of the LED wristband and is treated as hexadecimal characters. The color code is O, P, M, or B in order from class 1 to class 4 for VMT in Table 2, and the system lights the LED in the color corresponding to the input code. Fig. 12 shows a wireless communication system between the RMS main unit and the full-color LED light-emitting wristband. Wireless communication between devices is performed by the RMS main unit acting as the coordinator and the LED wristband acting as the end device. Also, if data on the wristband is difficult to reach, the refuge manager can use a router device for relaying. This enables the system to communicate with all victims.

6 QUESTIONNAIRE EVALUATION OF VIMS

At Kanagawa Institute of Technology Academy Festival and open campus, we conducted a questionnaire evaluation of the prototype system for the RMS which is the solution method of Phase 1, and the VMT which is the solution method of Phase 2. Chapter 6 describes the contents and results of the questionnaire evaluation regarding them.

6.1 Questionnaire Evaluation on Phase 1 Prototype

As a survey method, we introduced the RMS prototype and the outline of this research after distributing the questionnaire form. Questions in the Phase 1 prototype (RMS) correspond to questions Q1.1 to Q1.3 shown in Table 3. The questionnaire target was 38 individuals who visited the RMS exhibits. In this questionnaire, Question 1.1 (Q1.1) and Question 1.2 (Q1.2) are four choices. Also, we asked the respondents to describe and answer Question 1.3 (Q1.3) arbitrarily. For the results for Q1.1 and Q1.2, the number of answers to the question is treated as a score, and the result of them is shown as a column graph with each average score (the full score is 4.00 points) in Fig. 13.

In Q1.1 "Do you understand the outline of RMS?" 29 answered "Very good" and 8 answered "Good". It was found that 97% of the visitors understood the system. Also, the average score for this question was 3.74 points.

In addition, for Q1.2, "Do you feel like using this system when a large-scale disaster occurs?", 21 visitors responded "Very good", and 9 responded "Good". On the other hand, it was thought that about 20% of the visitors could not decide to use this system because there were 7 non-responders. We asked the reason for one visitor who answered "Poor" in this question. As a result, he said "Because RMS and battery unit are a bit heavy."

Table 2 Classification colors for VMT and handling method

Medical Questionnaire	Class (Color)
・ MQ1: Two or more items selected from fever, headache, abdominal pain (diarrhea), and nausea. ・ MQ1: A total of four or more items selected ・ MQ4: "Treatment required within 3 days" selected	1 (orange)
・ MQ2: "Inoperable without assistance" selected for two or more items	2 (pink)
・ MQ1: One, two or three items selected from the 7 items ・ MQ2: "Some actions require assistances" selected for two or more items (Accepting up to one caregiver) ・ MQ3: Pregnant woman (9 months ~) or a family with an infant ・ MQ4: "No treatment required within 3 days" selected ・ The person who should move to a private room according to judgment by the administrator based on the contents of the description column	3 (purple)
・ MQ3: Pregnant woman (~8 months) ・ Person who can stay in a large room according to decision by the administrator or medical staff based on the content of description	4 (blue)

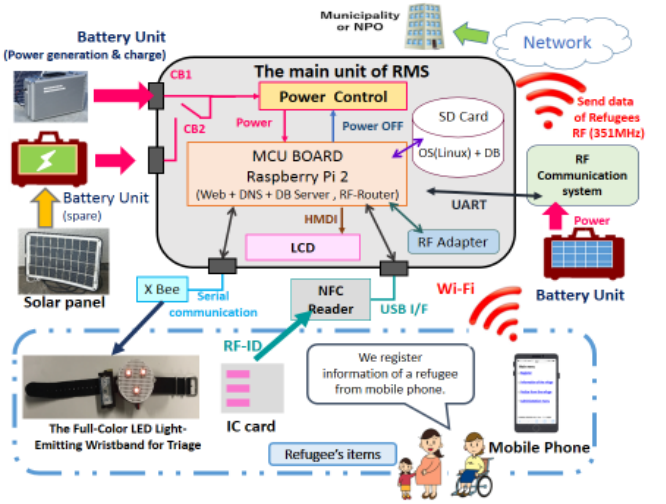


Fig. 7 RMS for Phase 2

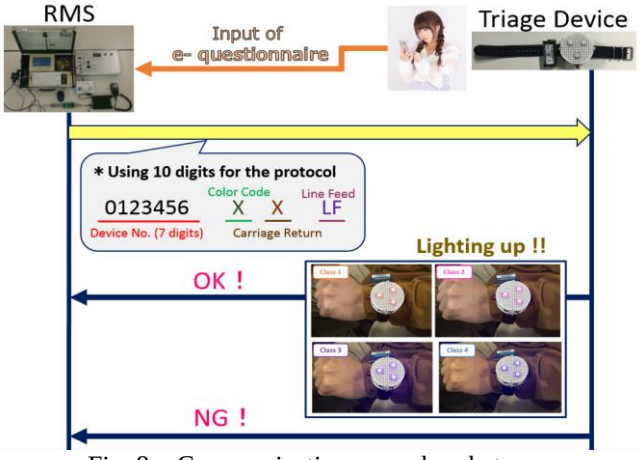


Fig. 8 Communication procedure between RMS and LED wristband

Table 3 Questionnaire questions on Phase 1 prototypes

	Contents
Q1.1	Did you understand the outline of RMS? 4 : Good 3 : Neither 2 : Poor 1 : Very poor
Q1.2	Do you feel like using this system when a large-scale disaster occurs? 4 : Good 3 : Neither 2 : Poor 1 : Very poor
Q1.3	Please tell us about opinions and requests about RMS.(descriptive expression)
Q2.1	Was it easy to input evacuation information and fill in *e-medical questionnaire form* on the RMS? 5 : Very good 4 : Good 3 : Neither 2 : Poor 1 : Very poor
Q2.2	Do you think that VMT is necessary after listening to the explanation of the electronic triage for persons requiring special care? 5 : Very good 4 : Good 3 : Neither 2 : Poor 1 : Very poor
Q2.3	Do you think the quantity of items in the VMT evacuation information and e-medical questionnaire form is appropriate? ・ Too much ・ Much ・ Appropriate ・ Less ・ Too less
Q2.4	What other kinds of people do you think need consideration other than the consideration items listed in "Decision Criteria" in Table 1 (of this paper)?

Furthermore, the following answers were obtained in response to the question Q1.3 “Please let me know if you have any feelings, opinions or requests about the refugee management system”.

- Pet information should also be included in the RMS.
- It is better to reflect facial pictures on the RMS.
- The system should also be available at a time other than when a disaster occurs.

6.2 Questionnaire Evaluation on Phase 2 Prototype

Questions in the Phase 2 prototype (VMT) correspond to questions Q2.1 to Q2.4 shown in Table 3. The questionnaire target was 20 individuals who visited the RMS exhibits. The answering method was to distribute the questionnaires asking questions Q2.1 to Q2.4 and to have the visitors fill in the questionnaire after the demonstration, as in the method described in the previous section. Questions 2.1 (Q2.1) to 2.3 (Q2.3) are five choices. Also, we asked the respondents to describe and answer Question 2.4 (Q2.4) arbitrarily. For the results, the number of answers to the question is treated as a score, and the result of them is shown as a column graph with each average score (the full score is 5.00 points) in Fig. 14(Q2.1&Q2.2) and shown as a pie chart Fig. 15(Q2.3).

As for the medical questionnaire, 10 visitors responded “Very good”, 6 visitors responded “Good”, 2 visitors responded “Neither”, and 2 visitors responded “Poor” about Q2.1 "Was it easy to input evacuation information and fill in *e-medical questionnaire form* on the RMS?". Also, the average score for Q2.1 was 4.20 points.

Next, about Q2.2 "Do you think that VMT is necessary after listening to the explanation of the electronic triage for persons requiring special care?”, 14 visitors responded “Very good”, 6 visitors responded “Good”. There was no one who answered "Neither", "Poor" and "Very poor". Also, the average score for Q2.2 was 4.70 points.

On the other hand, about 50% of the respondents answered “appropriate” to Q2.3 "Do you think the quantity of items of the VMT evacuation information and e-medical questionnaire form is appropriate?” The remaining 50% answered "too much" or "much".

As for Q2.4 “What other kinds of people do you think need consideration other than the consideration items listed in "Decision Criteria" in Table 1 (of this paper, section 5.1)”, there was an opinion that “children unattended by their parents also need consideration”.

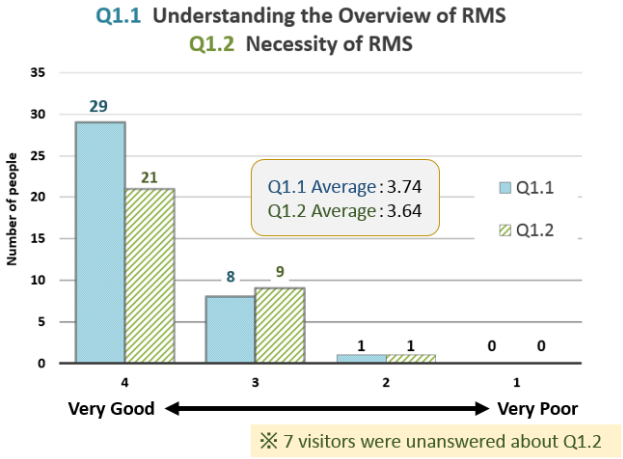


Fig. 13 Results of Questionnaire (Q1.1 & Q1.2)

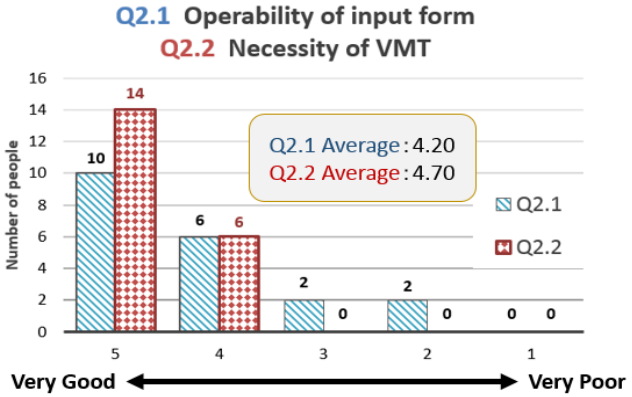


Fig. 14 Results of Questionnaire (Q2.1 & Q2.2)

Fig. 15 Results of Questionnaire (Q2.3)

6.3 Conclusion of questionnaire evaluation

As a result of conducting the questionnaire about the prototype system of Phase 1 and Phase 2, it is thought that many people highly appreciated the significance of the proposed VIMS system. Also, the average score was high in both Phase 1 and Phase 2. Above all, in the questionnaire on VMT in Phase 2, it was found that 100% of people understood the necessity of this research. On the other hand, as there were also suggestions regarding this system, we also felt that it is necessary to study and improve it.

7 CONCLUSION

This paper described the realization of the Victims Information Management System (VIMS), which centrally manages the disaster situation and the relief needs (including the need for relief for persons requiring special care) in refuge areas around the country. Since this VIMS is a large-scale system, we are conducting research in four phases, and promoting development toward the final goal of realizing the VIMS. Therefore, in this paper, we described in detail Phases 1 and 2 completed as prototype development.

In Chapter 4, we proposed an RMS corresponding to Phase 1, and a method for collecting and sending the relief needs in conjunction with a solar charging unit. We also proposed a system to manage entry and exit in refuges using UID management owned by the victims.

We proposed VMT for Phase 2 in Chapter 5. Specifically, we have proposed a system that mechanically determines the health status of victims from electronic questionnaires, and selects and guides the victims to evacuation spaces according to the patients' status.

In addition, a questionnaire survey on the prototype development of Phase 1 and Phase 2 of the VIMS research and development confirmed the effectiveness of this proposed system because many people answered that this proposed system is necessary.

Therefore, it is thought that Problems ① to ③ of the existing victim support systems mentioned in Chapter 2 can be solved by the contents of Phase 1 and Phase 2.

In the future, through Phase 3 and Phase 4, we plan to solve the issues ④ and ⑤ of the existing victims management system listed in Chapter 1.

REFERENCES

[1] Ministry of Internal Affairs and Communications of Japan : ”Survey result report on evacuation support of people who need assistance during disasters”, 2009. <http://www.shakyo.or.jp/research/2014_pdf/20140530_jisedai.pdf >

[2] Cabinet Office Government of Japan : ”Survey result report on promotion of comprehensive measures on evacuation”<http://www.bousai.go.jp/ kaigirep/ houkokusho/hinan_taisaku/houkoku.html> , (referred Sept.2016).

[3] Kosuke Akasaka, Keiichi Abe *et al.*, ”Deveropment and evalation of Refuge Management System using ICT at the time of Large scare disaster *(In Japanese)”*, Information Processing Society of Japan, CDS, Vol.7, No.3, pp.15 - 25, Sept.2017.

[4] Article 8, Paragraph 2 of the Disaster Countermeasures Basic Law of Japan, < http://www.japaneselawtranslation.go.jp/law/detail_main?re=01&vm=04&id=3092>, (referred Sept.2016).

[5] Information and Communication Technical Committee in Japan : “"Security guideline for ICT tool (cloud service) useful for large-scale disaster response” , First Edition,

[6] Kosuke Akasaka, Keiichi Abe *et al.*, ”Deveropment and evalation of Refuge Management System using ICT at the time of Large scare disaster *(In Japanese)”*, Information Processing Society of Japan, CDS, Vol.7, No.3, pp.15 - 25, Sept.2017.

[7] Kosuke Akasaka, Keiichi Abe *et al.*, “Proposal on Refuge Management System for Large Scale Disaster for Persons requiring special care by using triage “, IEEE BigComp2019 proceedings, pp.525-531, Feb. 2019(ISBN:978-1-5386-7789-6).

[8] Akari Hamamura, Egusa Nobuyuki *et al.*, : “AkariMap : Evacuation Support System for Everyday Use in Offline Environment (in Japanese)” , IPSJ DICOMO 2014,pp.2070-pp.2078, Jul.2014.

[9] Microsoft of Japan : “About the earthquake disaster recovery support system”, <>, (referred Sept. 2016).

[10] Nishinomiya City Urban Improvement Corporation : “Outline of evacuees support system” , <http://www.nishi.or.jp/homepage/n4c/hss/>

[11] Mariko Ohara , *et al.*, : “ Residents in the evacuation center at the occurrence of a disaster Efforts of people who need assistance triage ” , Japan Disaster Reconstruction Association Journal, Volume No. 10 Vol.6-No.1, pp. 19-24, Jun. 2014.

[12] Noboru Kiyama, Teruo Higashino *et al.*, : “An Advanced Electronic Triage System for Rescue Operations in Disaster *(In Japanese)*”, Information Processing Society of Japan, Vol.51-No.9, pp.1916–1926, Sep. 2010.

[13] Tammara Massey, “The Design of a Decentralized Electronic Triage System, ” 2006.<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1839501/>

Data Gathring Method of all Sensor Nodes in the Shortest Time for LPWA System

Mikiko Sode Tanaka
International college of Technology
Global information and Management
Kanazawa, Japan

***Abstract*— LPWA is mainly used to gather sensor data. Sensor data is usually gathered at fixed times, such as 0 minutes or 30 minutes and the data is transmit to the gateway at the same time. This is due to the characteristics of human beings. Because people tend to prefer regular times. Therefore, in LPWA, a lot of data is transmitted at the same time, causing a collision and causing a problem that data can't be acquired correctly. This is an LPWA-specific problem that differs from a random call at an unspecified time, such as a mobile phone. This problem is difficult to solve with LPWA, which is used by an unspecified number of users like sigfox. However, in a local network limited to specific applications, it is possible to improve by devising multi-hopping routing. To solve this problem, an algorithm that the gathering data in the shortest time is proposed to fully extract the capacity of the network and the provide data at the time desired by the users. In this paper, we give an example of field server for rice field and propose an algorithm to gather data in a short time. This method is effective for low power consumption. The comparison results with a conventional method using a simulation demonstrate that the proposed method can collect all of the data in the shortest amount of time. Moreover, it was confirmed that the power consumption is also lower than the conventional method. Furthermore, it was confirmed that the rate of increase in time necessary for the parent node to collect data, due to the increase in the number of sensor nodes, is lower than that of the other methods.**

Keywords - LoRa, Rice field, Data gathering

I. INTRODUCTION

In the LPWA network, the time is often overlapped for the gathering sensor data and for sending them to the gateway. As a result, the user may not be able to confirm the data at a desired time. This is due to the characteristics of human beings. This problem is difficult to solve with LPWA, which is used by an unspecified number of users like sigfox. Sigfox has a single carrier system in each country, and the subscriber (child node) transmits data directly to the base station (parent node). Figure 1 show the routing method of sigfox. There is no coordination among the child nodes. That is, the sensor data is directly transmitted to the parent node at a time when the child node likes. This is the same as the mobile phone format. The difference from mobile phones is that much of the data is sensor data, and in many cases data is transmitted at a predetermined time. This is due to human characteristics, and sensor data is acquired at a fixed time, and there is a tendency to check at the fixed time. Therefore, transmissions are not distributed as in mobile phones, but the sensor are transmitted at the same time. Thus, a collision occurs and the data can't be observed at the desired time.

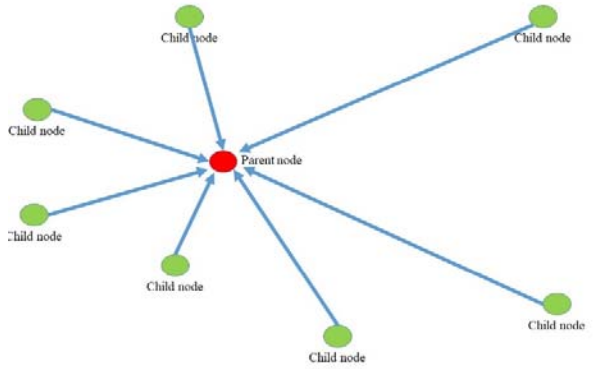


Figure 1. Routing method of sigfox

However, in a local network limited to specific applications, it is possible to improve by devising multi-hopping routing. For example, with Private LoRa, multi-hop is allowed, and you can freely set routing. To solve this problem, an algorithm that the gathering data in the shortest time is proposed to fully extract the capacity of the network and the provide data at the time desired by the users[1-3].

A chain-based protocol, Power Efficient Gathering in Sensor Information Systems (PEGASIS), has been proposed [4]. The child nodes are connected only to the closest chils node and are chained. Data is sent along the chain to the chain head (the last child machine of the chain) and from the chain head to the parent node. The chain head is randomly selected and has a duty to transmit data to the parent node. The time required to gather data for PEGASIS is long.

EPEGASIS [5] has been proposed as a method to solve the PEGASIS problem. The child nodes are broken down into clusters at a distance from the parent node, and a chain is created in each cluster in the same way as PEGASIS. Each cluster sends data to the chain head along the chain. When data gathers in the chain head in each cluster, the chain head transfers data to the parent node. Therefore, the time required for data collection can be reduced compared to PEGASIS.

CHIRON[6] has been proposed as a method for achieving further improvement. It is characterized by using Beam Star technology [7] to divide into smaller cluster areas than EPRGASIS. In each cluster, build a chain. Initially, the data for each cluster is collected by the chain head in the same way as PEGASIS. Next, a chain between chain heads is created in order from the chain head farthest from the parent node, along

which data is sent to the parent node. CHIRON can shorten the time required for data collection in each chain by shortening the chain length, but has the problem that the time required for gathering data in each area to the parent node increases.

In this paper, we give an example of field server for rice field and propose an algorithm to gather data in a short time with low power consumption. The proposed algorithm improve the gathering time problem. The results of a comparison with the conventional method using a simulation to demonstrate that the proposed method can collect all of the data in the shortest amount of time. Moreover, it was confirmed that the power consumption is also lower than the conventional method. Furthermore, it was confirmed that the rate of increase in time is necessary for the parent node to collect data due to the increase in the number of sensor nodes is lower than that of the other methods.

II. SENSOR WIRELESS NETWORK FOR RICE FIELD

In Japan, the average age of farmers has been on an upward trend, recently [8]. To reduce the burden on farmers, a rice cultivation field management system using a field server has been studied [9]. There are some commercial tools available [10, 11]; however, their introduction is not proceeding as expected. One reason is that the introduction and operating cost is expensive.

To reduce the introduction and operating cost, a system which uses LoRa [12] is proposed [13]. LoRa is a communication standard for IoT. It uses the ISM band; therefore, communication fees are not incurred. In addition, it also has features such as low power usage and a long transmission distance.

Based on our experiments, LoRa had the ability to conduct direct communication without a repeater in a 3 km × 3 km rice field of a rice cultivation agricultural corporation in Ishikawa Prefecture [11]. In addition, since repeaters are unnecessary, it is possible to reduce the installation cost. Moreover, we have confirmed that Wi-SUN and Zig-Bee do not have the ability to conduct direct communication without a repeater in a 3 km × 3 km rice field. Therefore, LoRa can be utilized as a communication standard for rice fields.

Installing a power supply is difficult in rice fields, unlike other fields. In addition, it is difficult to install a large-sized power generation device like a solar panel, as it is an obstacle to agricultural work, or to install wiring for a power supply inside the field. Therefore, the field server should operate on an electric battery at least during the rice cultivation period. Therefore, from the viewpoint of low power consumption, the adoption of LoRa is effective.

In fact, the communication speed of LoRa is lower than other communication methods such as Wi-Fi and Bluetooth. The real-time nature of data verification is important in a rice field [14]. Therefore, the data gathering speed is the most important requirement. In addition, a direct method is adopted in LoRaWAN. For this reason, the data gathering time is slow and the power consumption tends to increase.

III. SYSTEM CONFIGURATION OF RICE CULTIVATION MANAGEMENT SYSTEM

In the LPWA network, the time is often overlapped for the gathering sensor data and for sending them to the gateway. As a result, the user may not be able to confirm the data at a desired time. To solve this problem, an algorithm that the gathering data in the shortest time is needed to fully extract the capacity of the network and the provide data at the time desired by the users[8-11]

A rice cultivation management system monitors the air and water temperature of rice fields. Figure 2 shows the overall view of the rice cultivation management system. This system consists of a field subsystem, a home subsystem, and a management subsystem. The field subsystem is installed in the field, acquires air and water temperature data and transmits the acquired data to the home subsystem via a wireless network. The home subsystem is a system that uploads the data sent from the field subsystem to the management subsystem via the Internet. The management subsystem is a system that analyzes and displays the collected data and allows farmers to check the data accumulated on the cloud via the Internet.

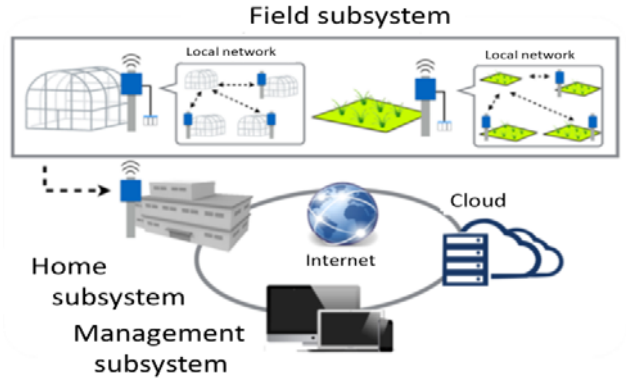


Figure 2. Target rice cultivation management system

In the field server subsystem, a plurality of child nodes are held; each child node acquires data once per hour, and transmits the acquired data to the home subsystem which is a parent node. This single process is called round one. Table 1 shows the frame format used for transmission. The frame is composed of a transmission destination, a transmission source, and a payload. Table 2 shows the sensor data transmission format stored in the payload of Table 1. In addition, the parent node can communicate with at most one child node at the same time, and the child node can communicate with at most one child node or parent node. In a single process, the power is turned off after system startup and after data acquisition, acquisition of a data transmission, and transmission are completed. The time required for this series of processes is defined as one round of operation time. This operation time is the time required to gather data to the parent node. In each child node, the data merging processing is performed as necessary. By this merging process, it is possible to shorten the time required for transmission.

Synchronization between the parent node and child nodes is carried out by using the current time transmitted according to the transmission request sent from the parent node [13]. When the child node receives the current time, the time is updated to the current time.

TABLE I. FRAME FORMAT (BYTE LENGTH)

Destination	Sender	Payload
1	1	variable

TABLE II. PAYLOAD FOR SENSOR DATA (BYTE LENGTH)

Child ID	Sensor data 1	Sensor data 2	Sensor data 3	Sensor data 4	sensor data 5
2	2	2	2	2	2

IV. LOW DELAY DATA GATHERING ALGORITHM

The direct method is adopted in LoRaWAN. However, since this method has a long data-gathering-time problem, in this paper we propose a method that has a low power consumption and a short data-gathering time. The data collection algorithm is shown below.

① Turn on the power of each child node and start them up.

② Send a transmission request by broadcasting from the parent node to the child nodes.

③ The child nodes measure the sensor data.

④ The child node closest to the parent node transmits the data to the master unit. The remaining child nodes establish connection with each other with the shortest distance and transmit the data from the remote child node to the pair child nodes.

⑤ Turn off the power of the child nodes that have completed transmitting the data.

⑥ Repeat the process of ④, ⑤ until all child nodes transmit data and their power has been switched off .

The number of steps T required for data gathering can be expressed by the following equation, where n is the number of nodes:

$$T = \lceil log_2 n \rceil \quad (1)$$

V. TRANSMISSION NETWORK GRAPH

We describe the procedure for creating a transfer network graph. A transmission network graph is a graph that describes the order, direction, and transmission timing of data transmission between a child node and a child node, or between a parent node and a child node. Data collection is performed according to this graph. The transfer network is defined by the effective graph G = (V, E). V is a set of the parent node and the child nodes. E is a set of edge, and edges e = {i, j} indicate that data is transmitted from node i to node j. E_m is set of edges which data is sent in m steps.

Here, in order to simplify the following discussion, the symbols are defined as follows.

V: {N1, N2} node set
N1: set of parent node
N2: set of child nodes
n_i: parent node
n_j: child node j where j=1, 2, 3, ...
l: the number of parent node
m: the number of child node
e=$\{n_j, n_i\}$:transmission from n_i to node n_j
E_m: set of edges in the step m
E=$\{E_1, E_2, ...\}$: set of edges
F: set of transmission completion nodes
E: set of un-transmission completion nodes
Nearest(j, G): The node with the closest Euclidean distance to node j in node set G for which transmission has not been completed
Farthest(j, G): The node with the farthest Euclidean distance to node j in node set G for which transmission has not been completed
t: step number
k: tradeoff variable of processing completion time and power consumption

The transfer network graph creation flow is shown below.

```
Procedure Routing(V， E)
1: F={n_0};
2: G=N2;
3: t=0;
4: k=4;   /* Processing time and power trade-off factor*/
5: While( There are k or more nodes belonging to G) {

6:     /* Transmission to the parent node*/
7:     t=t+1;
8:     i= Nearest(n_0, G);
9:     /* Find the nearest child node to the parent.  */
10:    Create branch {n_0, n_i} and put in set E_t representing processing at t time step
11:    Delete node i from un-transmit set G.
12:    Put the node i in the transmission complete set F.

13:    /* Data transmission between child nodes */
14:    H=G;
15:    While(There is a node that belongs to H) {
16:        j=Farthest(n_0， H);
17:        Delate j from H.
18:        i=Nearest(j， T).
19:        Delate i from H.
20:        Generate edges {i, j} and put them in the set E_t representing the processing at t time steps
21:        Delete node j from un-transmit set G.
22:        Put the node j in the transmission complete set F.
23:    }
```

```
24:}
25: While(There is a node that belongs to G) {
26:     /* Transmission to the parent node */
27:     t=t+1;
28:     i= Nearest(n_0, G);
29:    /* Find the closest child node to the parent node. */
30:     Generate edges {n_i, j} and put them in the set E_t
representing the processing at t time steps.
31:     Delete node j from un-transmit set G.
32:     Put the node j in the transmission complete set F.
33:}
```

The movement of the flow is explained using the example of one parent node and 11 child nodes shown in Fig. 2. In the first step, the child nodes 5 nearest to the parent node is selected, and an edge representing transmission to the parent node is made. Next, the child node 11 which is farthest from the parent node is selected, and an edge from the child node 11 to the child node 7 is made. Similarly, an edge from child node 3 to child node 2, an edge from child node 10 to child node 9, an edge from child node 8 to child node 6, an edge from child node 1 to child node 4 are created, and the first step is complete.

In the second step, as in the first step, an edge from the parent node to the nearest child node 9 to the parent node is created, an edge from the child node 7 to the child node 6, an edge from the child node 2 to the child node 4 is created. In the third step, an edge from the child node 4 to the parent node, and in the fourth step, an edge from the child node 6 to the parent node are created.

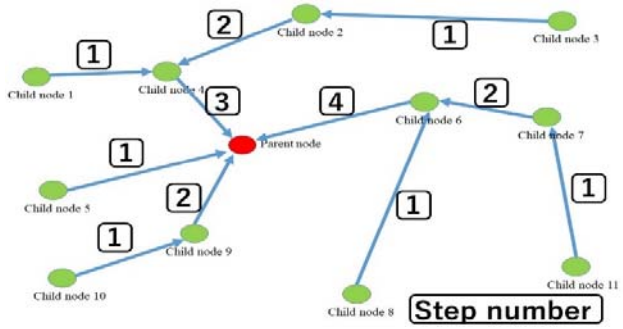


Figure 3. Operation example of one round

Next, the data gathering procedure will be described. Data is gathered to the parent node according to the tree structure of the graph created in the transfer network graph creation flow. Here, the operation is explained using the example in Figure 3.

In the first step, transmission corresponding to the edges belonging to the branch set E_1 representing the processing in the first step is performed. That is, transmissions from the child node 5 to the parent node, from the child node 11 to the child node 7, from the child node 3 to the child node 2, from the child node 10 to the child node 9, from the child node 8 to the child node 6, and from the child node 1 to the child node 4 are performed. The child node receiving the transmission merges the data held by itself and the transmitted data, creates a frame, and prepares for the next transmission. In the second step, transmission corresponding to the edge belonging to the edge set E_2 representing the processing in the second step is performed. That is, transmissions from the child node 9 to the parent node, from the child node 7 to the child node 6, and from the child node 2 to the child node 4 are performed. The child node receiving the transmission merges the data held by itself and the transmitted data, and creates a frame. In the third step, transmission corresponding to the edge belonging to the branch set E_3 that is transmission from the child node 4 to the parent node is performed. In the fourth step, transmission corresponding to the edge belonging to the branch set E_4 that is transmission from the child node 6 to the parent node is performed.

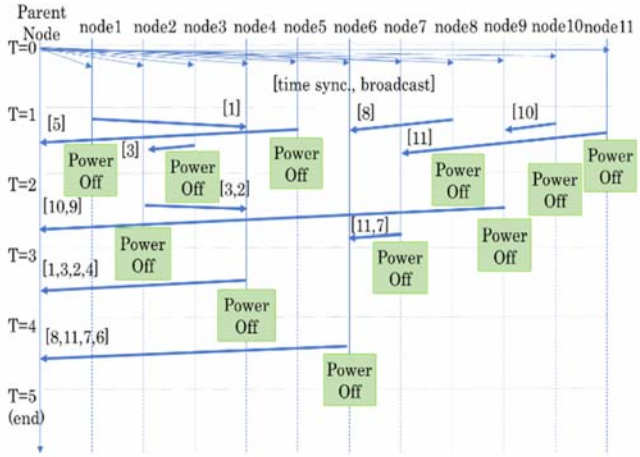


Figure 4. Operation example of one round

Figure 4 shows a sequence diagram of data exchange between the parent node and child nodes. The data of each child nodes is merged step by step, and all data is sent to the parent node in the fourth step.

The number of child nodes in operation at step T is defined as N (T). The number of steps required to gather data can be expressed by the following equation.

$$N(T+1)=\lceil (N(T)-1)/2 \rceil$$

Namely, T where N (T + 1) = 0 is a necessary step to transmit all data to the parent node. For example, in the example of Figure 3,

$$N(0)=10$$
$$N(1)=\lceil (10-1)/2 \rceil=5$$
$$N(2)=\lceil (5-1)/2 \rceil=2$$
$$N(3)=\lceil (2-1)/2 \rceil=1$$
$$N(4)=\lceil (1-1)/2 \rceil=0$$

In this example, since the number of child nodes is 10, N (0) = 10.Then, all data can be transmit to the parent node in 4 steps.

VI. PERFORMANCE EVALUATION BY SIMULATION

A comparison was made by simulation to check the data gathering time. Here, we focus on the prompt confirmation of the condition of the rice field required for rice cultivation. Therefore, we verify the data gathering time and the power consumption required to collect all the child node data. Being able to verify the data within an allowable time is an important item even if the number of child nodes is increased. The simulator was written in C language. First, the method of calculating the power consumption will be described. Owing to the fact that the parent node is always supplied with power, its power consumption is excluded from the scope of consideration and only child nodes are calculated. Each child node has a battery and has a capacity of 75,000 mWh. The child node has six modes (Table 3) according to the operating condition. The current consumption was set from the actual power consumption measurement report [13] using LoRa. The transmission power depends on the distance. Operation voltage is set at 2.5 V. The time of data merge is not considered because it is 1/100 or less of the time required for transmission and reception. Also, since the data transfer network is created only once at first, the influence on the power consumption is not considered because it is small.

TABLE III. SIX MODE OF CHILD NODE

State	State name	Power consumption (mA/s)	Processing time (s)
1	Power OFF	0	-
2	Data transmission mode	53, 62, 69, 78	3.4
3	Data reception mode	13.5	3.4
4	System startup mode	Parent 78 Child 13.5	3.4
5	Sensor data acquisition mode	60.0	60.0
6	Standby mode	2.7	3.4

The sensor data to be acquired is the data on temperature and humidity. In each round, when each child node reaches the activation time, the system is started and the data is acquired from the sensor. Next, the parent node transmits a data transmission request to all of the child nodes. Upon receiving the data transmission request, the child node measures the data from the sensor, and transmits the data to the parent node. Each child node turns off the power source. In this simulation, it is assumed that the time synchronization is perfectly performed. It was assumed that there was no failure of data transmission / reception. In fact, the merging time of the data by the child node was ignored in this simulation. Given that the payload can be transmitted at a data rate of up to 60 bytes, the transmission time can be shortened by the data merging process.

Table 4 shows the three instances of test data created. The child nodes have been randomly assigned within the circle of a specified radius, centered on the parent machine.

TABLE IV. DETAILS OF SIMULATION DATA

	#Parent	#Child	Radius(m)
DATA1	1	99	500
DATA2	1	200	500
DATA3	1	300	3000

Comparisons were made between the following methods: Direct [1], PEGASIS [4], EPEGASIS [5], CHIRON [6] and the proposed method. In addition, the following three items are compared to enable the farmer to confirm the data of the rice field in real time, while reducing the data gathering time and operating with batteries from the time of rice planting to harvesting.

i. Time required for data gathering in one round.
ii. Total power consumption of all child nodes in one round.
iii. Maximum working days (the number of days during which all the child nodes are in operation).

The simulation results of DATA 1 are shown in Table 5. Table 5 shows the operation time, which is the time required for the parent node to gather data, the total value of the power consumed by the child node, and the number of operable days. Figure 5 shows the size of data received by the parent node in the one round. The vertical axis in Figure 5 represents the data size (byte) received by the parent node, and the horizontal axis represents operating time (seconds). Figure 6 shows the number of child node whose transmission processing has been completed and the power has been turned off within one round. The horizontal axis shows the operating time (seconds), and the vertical axis shows the number of child nodes whose power has been turned off. It can be confirmed that the slope of PEGASIS is the slowest.

Compared to other methods, the proposed method shows that the time to gather data to the parent node is the shortest. Moreover, the total power consumption is the smallest, based on the simulation, it can be concluded that it can operate for a period that extends from rice planting to harvesting.

TABLE V. SIMULATION RESULTS OF DATA1

	Data collection time / Round (s)	Total power consumption of child nodes /Round (mAS)	Number of working days (Day)
Direct	400.0	74508.5	372
PEGASIS	1902.9	299237.6	58
EPEGASIS	226.6	55253.0	384
CHIRON	158.6	40131.1	448
Proposed	**114.4**	**34177.0**	**566**

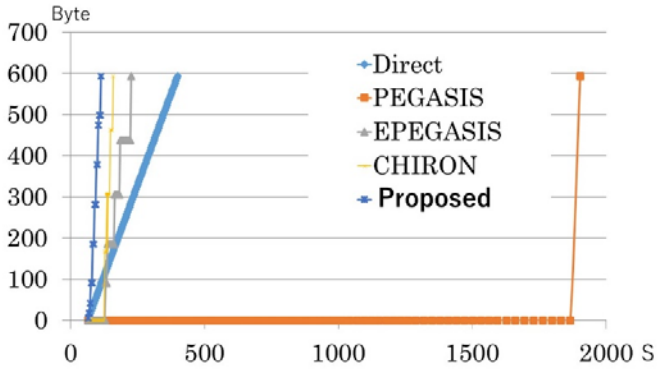


Figure 5 Number of data received by parent node of DATA1.

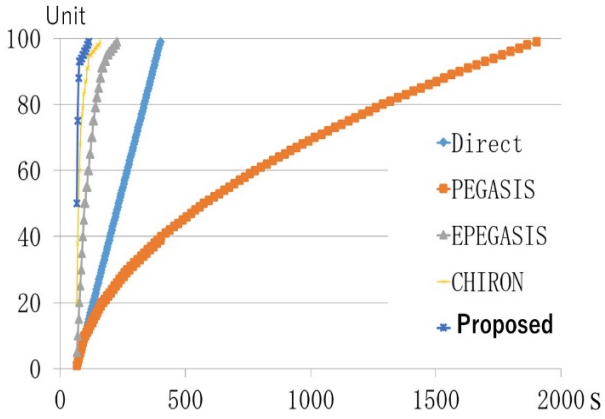


Figure 6 Number of power off nodes of DATA1.

TABLE VI. SIMULATION RESULTS OF DATA2

	Data collection time / Round (s)	Total power consumption of child nodes /Round (mAS)	Number of working days (Day)
Direct	743.4	243604.6	210
PEGASIS	7135.4	1763010.2	18
EPEGASIS	721.0	201747.2	195
CHIRON	248.7	90308.8	209
Proposed	**121.2**	**70925.4**	**231**

The simulation results for DATA 2 are shown in Table 6. Compared to other methods, the proposed method shows that the time for the parent node to collect data is the shortest. In addition, it is understood that the total power consumption is the smallest, and the lifetime is the longest. In the simulation, we can see that it can operate for a period that extends from rice planting to harvesting. In addition, although the proposed method doubles the number of child nodes as compared with DATA 1, it can be seen that the increase in time required for data gathering is as short as about 7 s.

The simulation results for DATA 3 are shown in Table 7. Compared to other methods, the proposed method shows that the time to for the parent node to collect data is the shortest. In addition, it is understood that the total power consumption is the smallest, and the lifetime is the longest. Moreover, in terms of the theoretical value, it can be understood that it can operate for a period extending from rice planting to harvesting. In addition, compared to DATA 1, the number of child nodes is tripled, but the increase of time increase necessary for collection is short as about 10 s.

TABLE VII. SIMULATION RESULTS OF DATA3

	Data collection time / Round (s)	Total power consumption of child nodes /Round (mAS)	Number of working days (Day)
Direct	1083.4	519160.0	67
PEGASIS	15771.0	5672543.4	8
EPEGASIS	695.8	487720.8	97
CHIRON	322.1	168756.0	151
Proposed	**124.6**	**117871.6**	**185**

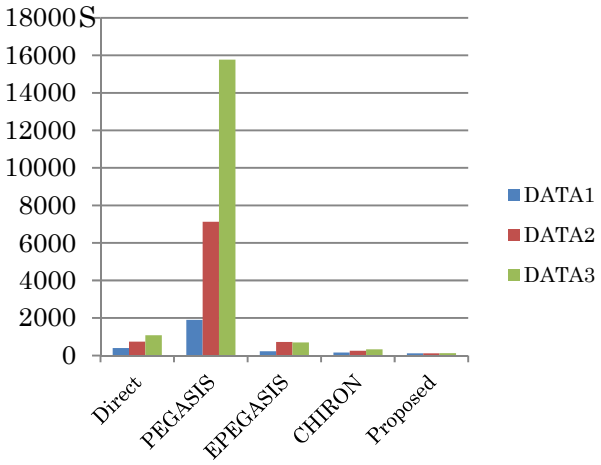


Figure 7. Comparison of time required to gather data in one round.

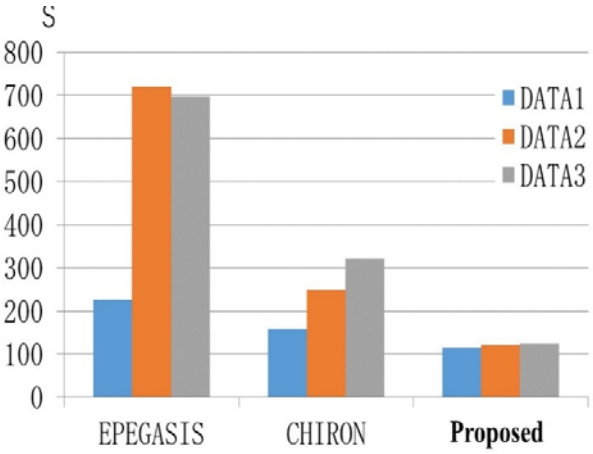


Fig. 8 Excerpt from the figure 7

Figure 7 displays the time to gather data to the parent node in one round. Figure 8 is excerpt from the figure 7. The vertical axis represents the operation time (seconds). For the Direct and PEGASIS methods, as the amount data increases, the time required to gather data to the parent node rapidly increases. For the proposed method, the rate of increase of the time required for data gathering according to an increase in the amount of data is the lowest, and data can be collected in a short time even if the number of child nodes increases. In addition, the elongation rate is about a 5% growth with respect to an increase in the number of child nodes. This result confirms the validity of Equation (1) and also indicates that the proposed method is superior to other methods in terms of the data gathering time.

VII. CONCLUSION

In this paper, we give an example of a field server for rice field and propose an algorithm to gather data in a short time, avoiding the collision caused by sensor data being acquired at the same time and transmitted simultaneously.

It is difficult to supply power to rice fields unlike other fields, so low power consumption is an important factor. In addition, the line usage fee required for communication must be reduced as much as possible from the standpoint of the entire cost required for agriculture. Therefore, we investigated interfiled communication using LoRa which is an IoT communication standard, low power consumption using the ISM band and a long transmission distance. LoRa has a low power consumption and a long transmission distance compared to other similarly used communication methods, but has the disadvantage of relatively slow communication speeds.

Therefore, to satisfy farmer requirements, we proposed a data gathering method that enables data to be collected in a short time between child nodes and a parent node. The proposed method has the feature of shortening the time required for data collection by performing the data merging processing and simultaneous transmission. As a result of the comparison using the simulation, we confirmed that the time required to gather child node data to the parent node is the shortest. Moreover, it was confirmed that the power consumption is smaller than that of conventional systems. Furthermore, we confirmed that the elongation rate of time necessary for gathering data to the parent node due to an increase in the number of child nodes is lower than that of the other methods.

Therefore, this proposal makes it possible to transmit the situation of a rice field in a timely manner and can be considered to be a useful method for reducing the burden of agricultural work.

REFERENCES

[1] Shalli Rani, Syed Hassan Ahmed, “Multi-hop Routing in Wireless Sensor networks” Springer.

[2] Sherin Abdel Hamid, Hossam S. Hassanein, Glen Takahara, “Routing for Wireless Multi-Hop Networks” Springer.

[3] Selmic, Rastko R., Phoha, Vir V., Serwadda, Abdul, “Wireless Sensor Networks Security, Coverage, and Localization,” Springer International Publishing.

[4] Lindsey, S., & Raghavendra, C. (2002). “Pegasis: power-efficient gathering in sensor information systems.” Proceedings of the Ieee aerospace conference proceedings, 2002. (pp. 3-1125 - 3-1130 vol.3). IEEE. 10.1109/AERO.2002.1035242.

[5] Jung, S.-M., Han, Y., & Chung, T.-M. (2007). “The concentric clustering scheme for efficient energy consumption in the pegasis”. Proceedings of the The 9th international conference on advanced communication technology (pp. 260-265). Gangwon-Do: IEEE. 10.1109/ICACT.2007.358351.

[6] Chen, K., Huang, J., & Hsiao, C., “Chiron: an energy-efficient chain-based hierarchical routing protocol in wireless sensor networks”. Proceedings of the Wireless telecommunications symposium, wts'2009 (pp. 183-187). USA: IEEE.

[7] Mao, S., & Hou, T. (2007). Beamstar: an edge-based approach to routing in wireless sensor networks. Transactions on Mobile Computing, 6(11), 1284-1296.

[8] Ministry of Agriculture “Statistics on agricultural labor force”, http://www.maff.go.jp/j/tokei/sihyo/DATA/08.html, (2016/11/02 reference).

[9] Kyokazu Kurosawa, Isamu Iizima, Yoshiki Amemiya, Shunya Yamamichi, Masaharu Toyota, Mikiko Sode Tanaka," Development of operational control system for rice cultivation," IEICE-CS 2016-12-IE-CS-AVM-BCT, 2016/12/8-9.

[10] Vegetalia, PaddyWatch, http://field-server.jp/paddywatch/, (2016/03/02 reference)

[11] PS Solutions, e-kakashi, https://www.e-kakashi.com/, (2016/03/02 reference).

[12] https://www.lora-alliance.org/ (2017/04/02 reference).

[13] Terada Keitaro, Masaharu Toyota, Hirata Tadaaki, Takafa Yuya, Keiko Matumoto, Mikiko Sode Tanaka, “Proposal of communication protocol for field management using LoRa,” Information Processing Society DICOMO2017 Symposium, pp. 1671-pp. 1678, 2017/6/28-30.

[14] Koichi Tanaka, Mikiko Sode, Masakatsu Nishigaki, Tadanori Mizuno, “Delayed minimum routing method of wireless network for rice cultivation management system,” Information Processing Society DICOMO2017 Symposium, pp. 550 - pp. 557, 2017/6/28-30.

Enhancing 3D Maps using RTK-GNSS to help improve the Position Solution in Urban Areas

Aastha Sahni†, Kyosuke Nakaaki†and Tomoya Kitani†

†Graduate School of Integrated Science and Technology, Shizuoka University, Japan
{aasthas333, k-nakaaki, t-kitani}@kitanilab.org

Abstract - To achieve centimeter level accuracy out of GNSS, direct (LOS: line-of-sight) signals from adequate number of satellites are required to calculate precise position solution. To improve position solution, 3D maps seem to be a promising approach to detect indirect (NLOS: Non-Line-of-Sight) signals, which degrade the accuracy of the position solution in urban areas. The major problem lies in the availability of accurate 3D Maps. In this paper, we have introduced a concept to generate and enhance 3D obstacles map using two datasets of Aerial LiDAR data and observation data received by the user's GNSS receiver. A basis of 3D obstacles map can be obtained by using LiDAR data. To enhance 3D obstacles map using observation data, reduction in observed C_0NR (Carrier to noise density ratio) that occurs when the buildings obstruct the line-of-sight between the orbital satellite and receiver, has been used in the paper to evaluate the presence of obstacles. To validate our detection of signals being LOS/NLOS, we calculated the position solution and found out that position solution significantly improved by 57% after removing noisy signals due to obstructions by buildings. Fish eye camera was used to cross validate whether the signals detected as NLOS were due to presence of obstacles such as buildings and trees.

Keywords: GNSS, LOS, 3D obstacles map, LiDAR, SNR, DGNSS

1 INTRODUCTION

In recent years, many applications have required highly precise positioning. Real Time Kinematics-Global Navigation Satellite System (RTK-GNSS) can unlock potential for a host of new positioning technologies and can help achieve accuracy up to centimeters [1]. The main idea behind GNSS is to measure direct distances between satellites and a user located on the surface of the earth or in the lower atmosphere. However, GNSS positioning performance in dense urban areas is severely degraded due to the obstruction and reflection of the signals by the surrounding buildings. Sometimes, both LOS and NLOS signals are received that tend to interfere with each other and produce multipath errors in positioning solution [2]. These multipath errors eliminate the number of received satellites and severely degrade the quality of the solution. Multipath errors can be suppressed through proper satellite selection methods [3–6] which guarantee positioning accuracy. Filter or carrier-smoothed positioning [7] algorithms can perform better, but these require sufficient number of good quality GNSS signals to maintain solution accuracy which is challenging in urban areas. Our research group has been developing a forecast system [8] that will predict which place and time is capable of precise positioning in future by detecting NLOS satellites using 3D obstacles map and will provide us with the score of the predicted accuracy and ability of that location to perform precise positioning. The major problem lies in the availability of 3D maps. If we will have accurate 3D information of buildings, trees and other obstacles in advance, it will help detect NLOS satellites precisely and help improve the position solution.

3D maps can be generated by using only LiDAR data that is available at government organizations [9]. However, LiDAR data alone cannot be trusted because it can be old, not updated and may not contain the obstacles which are invisible through naked eyes. Although GNSS signals and their C_0NR or SNR (Signal-to-noise-ratio) alone can be exploited to reconstruct 3D urban scenes [10], it is not feasible for covering larger areas. Hence, we have come up with hybridization of datasets to achieve accurate 3D obstacles map Aerial LiDAR dataset and Observation dataset. The proposed method in this paper will help enhance already available, existing 3D LiDAR data set using observed data by user's receiver so that enhanced and accurate 3D obstacles map can be fed into our forecast system. The proposed method seems feasible as enhancing existing 3D LiDAR data is not tiresome, uses reliable observation data user's receiver and adds accuracy to the system. In experimental results, we confirmed 96.1% times the presence of buildings using the proposed method.

The mentioned precision forecast system [8] will help improve reliability of RTK-GNSS in urban areas. In this paper, the proposed SNR (Signal-to-noise ratio) based enhancement of 3D obstacles map contributes in providing obstacles information for the forecast system using a low cost rover. Moreover, the proposed SNR-based signal detection algorithm will help improve the position solution in dense urban areas.

2 DIFFERENTIAL POSITIONING THEORY

RTK-GNSS is a differential and interferometric positioning system that provides a receiver's position of centimeter-level precision in real time using positioning satellites and ground stations [11]. In the single point positioning, a receiver's position is calculated with the measured pseudo-range between each satellite and the receiver. A pseudo-range is a low-resolution distance information (within a meter-level) which

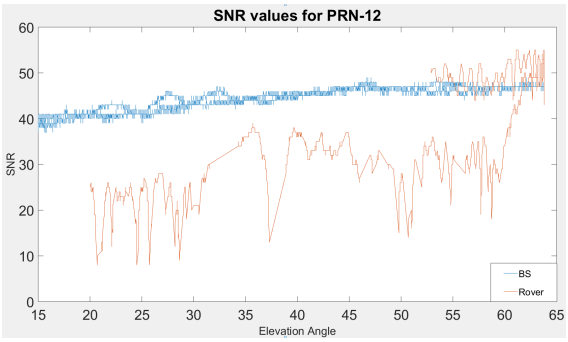


Figure 1: SNR Fluctuations in Multipath Environment

contains several errors due to atmospheric factors. On the other hand, in the kinematic RTK-GNSS positioning, a carrier phase is used instead of a pseudo-range as a higher-resolution distance information whose measurement resolution is within a few milli-meters. RTK uses a complicated algorithm to calculate the exact number of radio wavelengths between the satellites and the base station antenna, a process known as ambiguity resolution and yield either a fixed or float solution. In a fixed solution, the number of wavelengths is a whole number, or integer, and the algorithm is constrained to yield a whole number. The fixed solution signifies more accurate position and accuracy can be achieved up to centimeter levels. The accuracy in a Float Solution is not so high and can be of few meters. The main objective of differential GNSS is to set up a reference receiver at a known position, called Base Station, and estimating the errors. These estimates, called corrections, can be used in order to compensate for similar errors in another receiver, called rover, whose position is unknown. All rovers are relative to the base station and observe the same constellation of satellites at the same time. Hence, both, base station and rover record very similar errors and since, base station's position is known, rover's position can be determined accurately within few centimeters, and errors can be compensated easily.

2.1 Characteristics of SNR in Urban Areas

Carrier to Noise Ratio (C_0NR) is a measure of the received carrier strength relative to the strength of the received noise. Also, Signal to Noise Ratio (SNR) , is a measure of signal strength relative to background noise. C_0NR and SNR are often used interchangeably in GNSS world. We choose C_0NR or SNR over other observable types because multipath effects on SNR have the most unambiguous relationship to multipath and determined independently for individual satellites. In our experiment, one receiver was located on 112.49m high building in obstruction free horizon and another receiver was located on the ground surrounded by buildings and trees. The observations shown in Fig.1 shows the tendency of drop in SNR when the receiver is located in the multipath environment and SNR measurements are smoother when the receiver is in obstruction free environment.

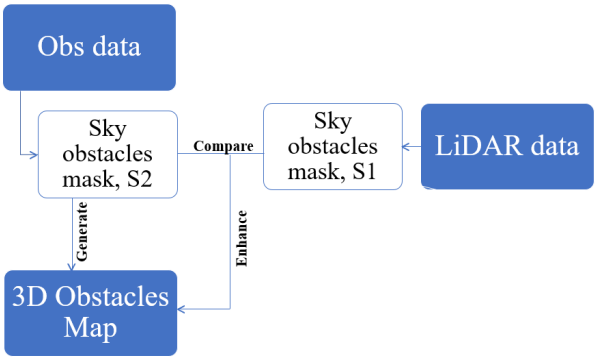


Figure 2: Approach towards 3D obstacles map

3 APPROACH TOWARDS 3D OBSTACLES MAP

To generate 3D obstacles map, observation data from the orbital satellites received by the user's receiver, such as SNR as mentioned in Section 2.1 and other observables, can be used for distinguishing which satellite signal is direct or indirect indicating the presence or absence of an obstacle. However, it is computationally time consuming and has drawbacks when it comes to covering larger areas. Hence, alternatively we will use this valuable observation dataset to enhance and improve existing cm-level precise Aerial LiDAR dataset as shown in Fig.2.

Aerial LIDAR has proven that it can provide three-dimensional (3D) information of the Earth surface with high accuracy. Despite that, only LiDAR Data cannot be trusted as it can be old, not updated and may not contain the obstacles which are invisible through naked eyes. Hence, two datasets are used to generate and enhance 3D obstacles map : Aerial 3D LiDAR dataset and Observation dataset measured by user's GNSS receiver. LiDAR data is usually archived in binary files in the LASer (LAS) format [12], which is an open standard of three-dimensional (3D) data storage [4]. The format, although efficient for the storage, is not particularly convenient for data management and dissemination. As the LAS standard does not allow for data cross-referencing, LiDAR dataset must be converted into understandable and organized format. Hence, we converted the point cloud data into 1-by-1 mesh grid [13] and came up with our own term "Sky Obstacles Mask" in the paper.

"Sky Obstacles Mask" refers to a buffer that contains data of an area, viewing towards the sky from a ground position, divided into two states of 0 and 1 at azimuth and elevation angles where:

0 = No Obstacle
1 = Obstacle

"Sky Obstacles Mask", S_{ldr}, for LiDAR dataset, and "Sky Obstacles Buffer", S_{obs}, for observation dataset, will be generated; S_{obs} will help enhance S_{ldr} and finally we will have enhanced 3D obstacles map.

3.1 Sky Obstacles Mask using LiDAR data

To generate sky obstacles mask, S_{ldr}, a point on a straight line, from the viewer towards the sky, is checked every 1m with the radius of 300m at every 10 degrees azimuth angle and every 5 degrees at elevation angle. If the point exists in LiDAR data, it signifies presence of some obstacle and hence, that point is specified as 1. If the point does not exist in point cloud data, it is specified as 0. We check it upto 360 degrees for azimuth angle and 90 degrees for elevation angle. Finally, we will get Sky Obstacles Mask, S_{ldr}. The JSON friendly format is mentioned below:

$$\{AZi : e_{i,0} \;\; e_{i,5} \;\; e_{i,10}.....e_{85} \;\; \}$$

where, AZ is azimuth angle from 0 to 360 degrees at every 10 degrees and e represents elevation angle from 0 to 90 degrees at every 5 degrees corresponding to azimuth angle. Or,

$$i \in \{0, 10, 20...350\} \quad and \quad j \in \{0, 5, 10...85\}$$

$$e_{i,j} \in \{0, 1\}$$

For example, sky obstacles mask at azimuth angle 0-10 degrees looks like {"AZ0":"1111111110000000000"}. This kind of format is used because it represents the presence or absence of obstacles based on azimuth and elevation angles at a certain location, unlike conventional elevation mask.

3.2 Sky Obstacles Buffer using Observation data

As raw LiDAR point cloud data is converted into understandable Sky Obstacles Mask, our next step is to enhance and modify this mask using observation data to achieve better accuracy. Referring to [14], it can be conferred that SNR of a signal from a satellite is highly dependent on the elevation angle of the satellite. Hence, Elevation angle-dependent SNR (Signal to Noise ratio) has been used from the observation data for every satellite to distinguish whether the signal is direct or indirect (0 or 1) . "Sky Obstacles Buffer" contains the location of the user receiver, azimuth angle, elevation angle and the state of 0 or 1 of the signal. Finally compare this buffer data with the LiDAR data obstacles mask at the same location for every angle segment, replace the old value in LIDAR data obstacles mask with observation data obstacles mask, enhance and improve it. Distinguishing a signal whether it is direct or indirect (0 or 1) is in itself a complicated process. Hence, the authors have narrowed down their research to predict whether the received signal will be specified as 0 or 1. The following sections in the paper discuss about the analysis of observation data.

4 PROPOSED METHOD FOR NLOS DETECTION

Based on the characteristic of SNR of the signal which says that the pattern of the signal fluctuates under static conditions,

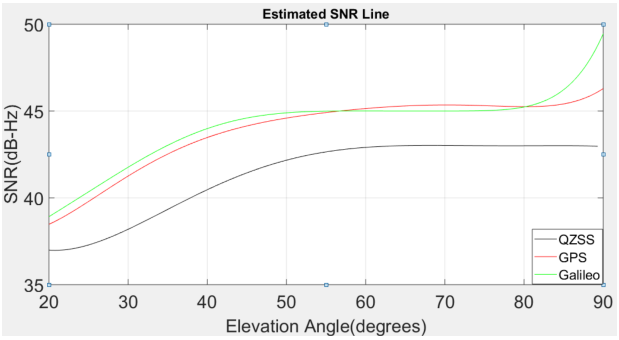


Figure 3: Elevation Dependent-Estimated SNR Line for different satellite systems

the differential SNR based distinction of signals has been used to detect NLOS signals. We used elevation dependent estimated SNR line as reference to set standard SNR values at corresponding angles in the obstacle free environment. The drop in SNR when the receiver is surrounded by obstacles is compared with the standard SNR depicted by the estimated line. Finally, we set a threshold value of SNR. If the observed SNR value exceeds threshold, the signal is specified as NLOS or LOS, otherwise. [5]

4.1 Elevation dependent estimated SNR line

We calculated the value of the SNR as a function of elevation in advance from a rover receiver and base station with antennas fixed in a multipath-free environment. Raw observation data of 24 hours is used to produce the SNR lines as shown in Fig.3. Then, we generated SNR line for GPS, QZSS and Galileo Satellite system by using Curve Fit Method (Linear Least Squares Method) [15, 16].

4.2 SNR Threshold

The difference between the observed and estimated SNR line is expressed as:

$$\text{A (t)} = SNR_{observed}(t) - SNR_{estimated}(elv)$$

where $SNR_{observed}$ is the observed SNR at t epoch (Epoch means the time interval between two consecutive time points) and $SNR_{estimated(elv)}$ is the estimated elevation-dependent SNR. Here, we used the estimated line for deriving SNR value instead of exact SNR value for a particular day because it is less time consuming and we do not have to collect data every time in open sky. We used fish eye camera to generate elevation mask using RTKLIB software [17] as shown in Fig.4.

This elevation mask includes azimuth and elevation angles of the boundaries of buildings. We used this elevation mask to compare and inspect the predicted state of the signal at various values of drop in SNR and hence, inspecting the predicted presence of building or not. SNR threshold value was determined statistically to predict the signal as 0 or 1 (LOS or NLOS) . We set the threshold value as 4 db-Hz as shown in Fig.5, where positive case implies NLOS and negative case

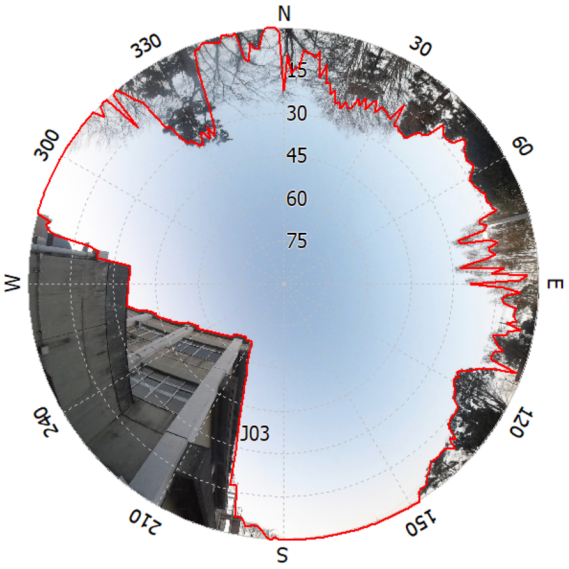


Figure 4: Elevation Mask

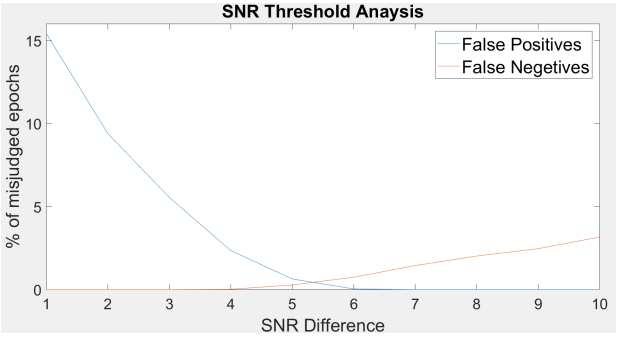


Figure 5: SNR Threshold Analysis

implies LOS. This threshold value was set because we are more concerned about false negatives i.e originally the signal is NLOS, but it is predicted as LOS, which will severely degrade the position solution. Hence, the less the false negatives, the better results are expected. If the SNR drops below this threshold value, the signal will be specified as 1 (NLOS) in Sky Obstacles Mask or 0, otherwise.

The flow chart of generating Sky Obstacles Mask using observation data is shown in Fig.6.

5 EXPERIMENTAL SETUP AND RESULTS

To test the performance of the generated sky obstacles mask, an experiment was conducted in our university. Sky Obstacles Mask was generated using Aerial LiDAR point cloud data as shown in Fig.7 based on the procedure described in Section 3.1 which signifies the available area for precise positioning solution. It was visually compared with the Google Maps and seemed to work fine. To generate Sky Obstacles Mask from observation data, we need a buffer that contains LOS or NLOS data for each satellite at every azimuth and elevation angle. Hence, differential SNR-based proposed method

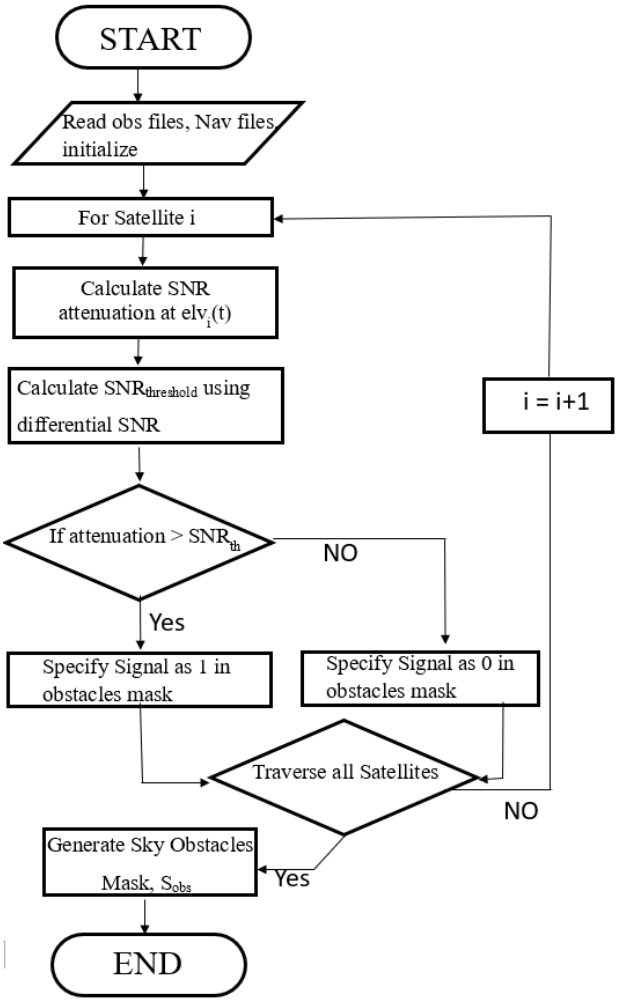


Figure 6: Flow chart for generating Sky Obstacles Mask, Sobs

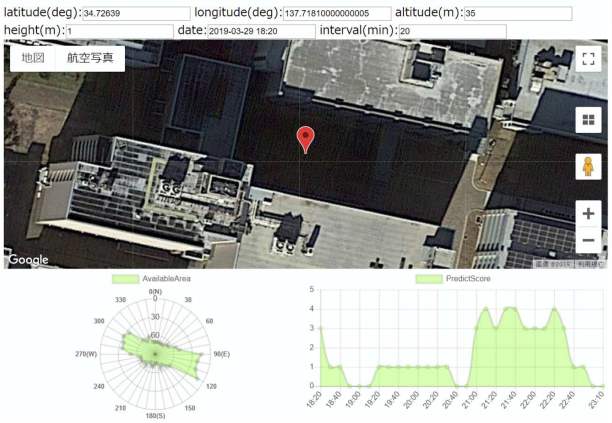


Figure 7: Sky Obstacles Mask using LiDAR data

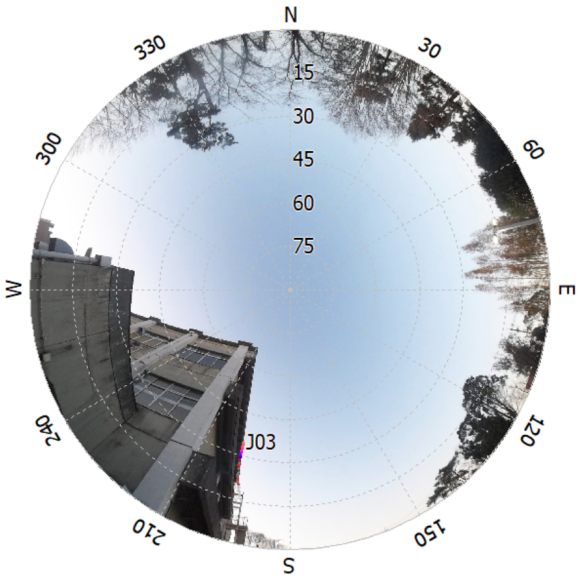


Figure 8: Experiment site

for LOS and NLOS distinction algorithm was tested using GPS data collected using a U-blox evaluation kit (EVK-M8T) which is a multi-constellation GNSS receiver.

The test data was collected in Shizuoka University on 22nd April 2019 as shown in Fig.8. The Base Station was located on 112.49m high building in obstruction free horizon and another receiver was located on the tripod stand at 1m above the ground surrounded by a medium height building. In addition to the interference of the surrounding environment, the signal strength is also affected by other factors such as antenna, atmosphere, ionosphere, etc. To remove the influence of other factors, the standard signal is measured by the same equipment in the open sky.

To cross validate our estimation of detecting LOS and NLOS signals, we used fish eye camera to compare those satellite signals with the estimated results of our proposed method. It was found that 96.1% of distinction results given by our method matched with the distinction results given by fish eye

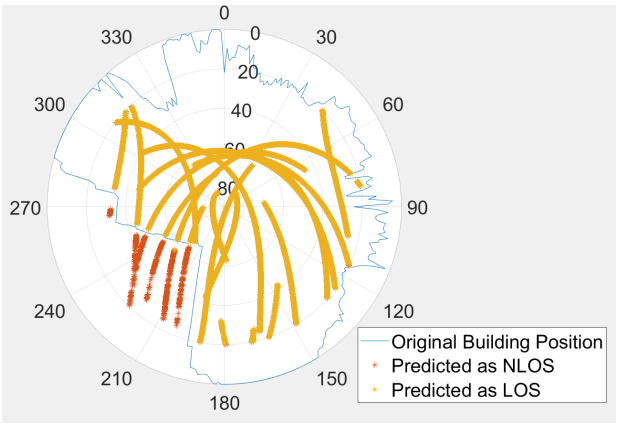


Figure 9: Predicted Signals w.r.t buildings' boundary

camera (Fig.9) .

The position solution was calculated using RTKLIB Software version 2.4.3 [17] to cross examine the fact that the detected signals were noisy and should be eliminated from the calculation of position solution. Table 1 shows results of position solution calculated at five-time intervals of 30 minutes each.

Table 1: Position Solution results comparison

Solution Stats		Before	After
T1	FIX Rate	14.3%	63.4%
	FLOAT Rate	85.7%	36.6%
	RMSE (m)	$H = 0.3181$ $V = 0.3359$	$H = 0.3082$ $V = 0.3212$
T2	FIX Rate	26.2%	85%
	FLOAT Rate	73.8%	15%
	RMSE (m)	$H = 0.8595$ $V = 0.6810$	$H = 0.1949$ $V = 0.3812$
T3	FIX Rate	31.3%	93.1%
	FLOAT Rate	68.7%	6.9%
	RMSE (m)	$H = 0.5951$ $V = 1.1473$	$H = 0.3404$ $V = 0.0841$
T4	FIX Rate	68.2%	96.3%
	FLOAT Rate	31.8%	3.7%
	RMSE (m)	$H = 0.1965$ $V = 0.1891$	$H = 0.0610$ $V = 0.1296$
T5	FIX Rate	77.2%	98.8%
	FLOAT Rate	22.8%	1.2%
	RMSE (m)	$H = 0.2483$ $V = 0.3449$	$H = 0.0087$ $V = 0.0078$

Table 1 shows that fix solution significantly improved after removing multipath signals. Hence, we can signify those signals as NLOS in our Sky Obstacles Mask.

6 CONCLUSION

To enhance 3D obstacles map, we used two sets of data: Aerial LiDAR data and observation data. The data was con-

verted to more understandable form, Sky Obstacles Mask. Sky Obstacles Mask generated using LiDAR data seems to work fine in predicting the available area to perform precise positioning solution. To generate Sky Obstacles Buffer using observation data, distinction between direct and indirect signal is an important step which in itself is a complicated subject of matter. We used differential SNR-based method to detect LOS and NLOS signals.

Experimental results show significant improvement in the position solution when the noisy satellites were detected and eliminated from the position calculation. This signifies that we can specify those multipath signals as NLOS in our Sky Obstacles Mask. To cross validate the results, we used fish eye camera to judge whether there was any obstacle when the signals became noisy or not. We found that 96.1% of predicted distinction results matched with the results given by fish eye camera.

We have shown that a standard GNSS receiver can detect NLOS satellites by measuring reduction in differential SNR. The SNR, elevation angles, azimuth angles and the distinction state of 0 (LOS) or 1 (NLOS) will be saved in the buffer that will help to detect and localize buildings with the help of generating Sky Obstacles Mask and finally use this mask to enhance and update existing LiDAR data.

For future work, distinction between LOS and NLOS signals can be processed by using machine learning. This will faster the process of enhancing 3D maps using observation data. Moreover, the user will be able to acquire its accurate position by eliminating NLOS signals provided by machine learning algorithms. The process to distinguish LOS and NLOS signals needs more research and leaves a scope for improvement.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP17H01731.

REFERENCES

[1] Y. Feng and J. Wang, "GPS RTK performance characteristics and analysis," *Journal of Global Positioning Systems*, vol. 7, no. 1, pp. 1–8, 2008.

[2] I. Ballester-Gurpide, E. Herráiz-Monseco, A. Miguel, M. Romay-Merino, and T. Beech, "Future GNSS constellation performances inside urban environments," in *Proc. of the 13th International Technical Meeting of the Satellite Division of The Institute of Navigation*, pp. 2436–2445, 2000.

[3] Y. Fang, Y. Hong, O. Zhou, W. Liang, and W. Liu, "A GNSS satellite selection method based on SNR fluctuation in multipath environments," *Int. J. Control Autom*, vol. 811, pp. 313–324, 2015.

[4] M. Adjrad and P. D. Groves, "Enhancing conventional GNSS positioning with 3D mapping without accurate prior knowledge," pp. 2397–2409, The Institute of Navigation, 2015.

[5] H. Tokura and N. Kubo, "Effective satellite selection methods for RTK-GNSS NLOS exclusion in dense urban environments," in *Proc. of the 29th International Technical Meeting of the Satellite Division of the Institute of Navigation, Oregon*, pp. 304–312, 2016.

[6] P. D. Groves and M. Adjrad, "Likelihood-based GNSS positioning using LOS/NLOS predictions from 3D mapping and pseudoranges," *GPS Solutions*, vol. 21, no. 4, pp. 1805–1816, 2017.

[7] T. J. Ford and J. Hamilton, "A New Positioning Filter: Phase Smoothing in the Position Domain," *The Institute of Navigation*, vol. 50, pp. 65–78, 2003.

[8] K. Nakaaki and T. Kitani, "An evaluation result of an accuracy scheme for RTK-GNSS Precision Forecasting," in *GPS/GNSS SYMPOSIUM 2017*, p. 202, 2017 (in Japanese).

[9] R. Idris, Z. Abd Latif, J. Abdul Hamid, J. Jaafar, and M. Ahmad, "Integrating airborne LiDAR dataset and photographic images towards the construction of 3D building model," *IOP Conference Series: Earth and Environmental Science*, vol. 18, 2014.

[10] K. Kim, J. Summet, T. Starner, D. Ashbrook, M. Kapade, and I. Essa, "Localization and 3D reconstruction of urban scenes using GPS," in *Proc. of 12th IEEE International Symposium on Wearable Computers*, pp. 11–14, IEEE, 2008.

[11] T. Kitani, H. Hatano, M. Fujii, A. Ito, and Y. Watanabe, "A cooperative GPS/GNSS positioning method with neighboring receivers," in *Proc. of International Workshop on Informatics (IWIN2014)*, pp. 145–153, Informatics Society, 2014.

[12] *LAS SPECIFICATION VERSION 1.4*. The American Society for Photogrammetry Remote Sensing, 2013.

[13] M. Boussaha, B. Vallet, and P. Rives, "Large scale textured mesh reconstruction from mobile mapping images and LIDAR scans," *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. IV-2, pp. 49–56, 2018.

[14] J. Hefty, "Multipath detection with the combination of SNR measurements Example from urban environment," *Geodesy and Cartography*, vol. 66, no. 2, pp. 305–315, 2017.

[15] B. V. Pai, W. A. W. Abidin, A. Othman, H. Zen, and T. Masri, "Characteristics of mobile satellite L-band signal in mid-latitude region: GPS approach," 2011.

[16] S. Deep, S. Raghavendra, and B. Bharath, "GPS SNR prediction in urban environment," *The Egyptian Journal of Remote Sensing and Space Science*, vol. 21, no. 1, pp. 83–85, 2018.

[17] T. Takasu and A. Yasuda, "Development of the low-cost RTK-GPS receiver with an open source program package RTKLIB," in *Proc. of International symposium on GPS/GNSS*, pp. 4–6, 2009.

Examination of Incorrect Image Selection Method for Image based Authentication Using Browsing History of Web Pages

Yusuke Iizawa†, Yoshitaka Nakamura‡, and Hiroshi Inamura‡

†Graduate School of Systems Information Science, Future University Hakodate, Japan
‡School of Systems Information Science, Future University Hakodate, Japan
{g2118002, y-nakamr, inamura}@fun.ac.jp

Abstract - In recent years, with the increase in the frequency of use of smartphones,terminals require high security user authentication. In current personal authentication of smartphone, “What you know” and “What you are” are generally used. However, they have problems with memory load, operation load, and use of user-specific persistent data such as a fingerprint.

In this paper, we propose a new personal authentication method to reduce memory load, operation load without using user-specific persistent data. As an approach of this method, we consider combining image based authentication with excellent storage load and operation load and behavior history based authentication using continuous information with change. The proposed method uses screen shot of the web page that the user most closely watched as correct image, and screen shots of web page not seen by user as incorrect images. We examined two methods, selection method excluding web pages of the same domain as correct image and selection method based on semantic distance with search keyword as selection methods for incorrect images. We also verified its effectiveness by comparing the two methods. As a result, it was shown that the incorrect image selection method based on the semantic distance to the search keyword is effective.

Keywords: Smartphone, Personal authentication, Web browser history, Image-based authentication, Screenshot

1 INTRODUCTION

In recent years, smartphones have become widespread, and many people use them on a daily basis.

Due to the characteristics of smartphones, users may operate smartphones anywhere they can connect to the network. Therefore, the opportunity to use the smartphone terminal in the public place where many unspecified people exist increases, and the risk that the data in the terminal is accessed due to the theft of the terminal itself is also increased. Since the internal data of the terminal contains a large amount of personal information, there is a risk that the theft of the terminal may cause unauthorized withdrawal from the online bank account or leakage of the personal information. Therefore, in order to ensure security of data in the terminal even if the terminal is stolen, it is important to provide a user authentication method with high security when using the terminal.

As a personal authentication method for smartphones, two are generally used: “What you know” and “What you are”.

“What you know” refers to an authentication method in which the user arbitrarily sets and uses secret information such as PIN (Personal Identification Number), character string password, pattern and so on. This authentication method needs to set secret information that is complex enough not to be guessed by others. Although this improves security against attacks based on user behavior observation such as peep attacks and guess attacks, it requires increased memory load of the user and complicated authentication operations. On the other hand, setting simple secret information reduces the user's memory load, but makes the authentication vulnerable to the aforementioned attack methods. In addition, since there is a tendency to simply set the password of the PC[1], it is considered that the same tendency is also made to the smartphone.

“What you are” is called biometric authentication. It refers to an authentication method that sets physical features such as fingerprints, irises and faces as secret information.Fingerprint authentication method is installed in many smartphones because the input operation of secret information is very convenient with only the minimum operation of “touch”. On the other hand, there is a problem of spoofing by using permanent data of user's biometric data [3] [4].

Because of these problems, a secure authentication method is required without using permanent biometric data.

2 RELATED WORK

Image based authentication is a personal authentication method that can reduce memory load and operation load. Image based authentication presents a group of images in which a correct image set as secret information by the user and incorrect images are mixed, and performs personal authentication based on whether the user can select a correct image. The basic steps of image based authentication are shown in the Fig.1. Images have the following effects on human memory[5].

1. Easy to memorize in large quantities compared to text data
2. Easy to memorize for a long time compared to text data

The fact that the memory of an image is better than that of a text, as described above, is called the “Picture Superiority Effect(PSE)” [6]. Because of this effect, image based authentication is said to be superior to “What you know” in terms of memory load. “Dejavu”[7], which uses images of geometric patterns, and “Awase-E”[8], which uses photos taken with mobile terminals as images, have been devised. However, in

order to perform image authentication, it is necessary for the user to collect a large number of images and register them in advance, which places a heavy burden on the user.

As a personal authentication method using personal data other than biometric information, there are some researches that acquire the user's behavior history according to sensor data obtained by a smartphone and perform implicit personal authentication based on this information[9][10]. However, behavior based authentication may be highly dependent on the surrounding environment.

Therefore some researches were conducted to identify the user's subjective experience information such as yesterday's breakfast and the visited shops, and to identify individuals.

Authentication using the user's mail reception history[11] and "PassFrame" using video recorded from the user's viewpoint have been proposed[12].

But, behavior history is easily associated with information about the user's life and relationships, so it is necessary to use an appropriate behavior history that is not associated with such information.

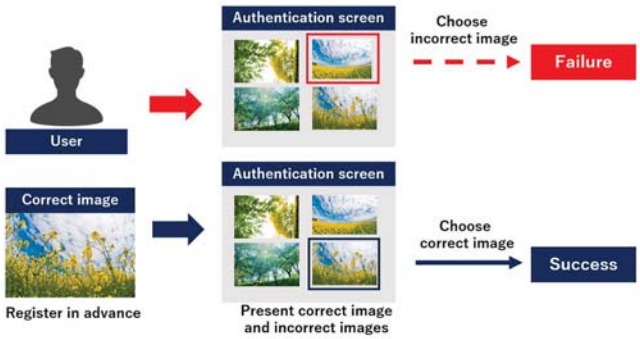


Figure 1: Basic steps of image authentication

3 Image based authentication using behavior history

3.1 Approach

By combining image based authentication and behavior history based authentication, the advantages of both methods can be obtained.

Web page is used as authentication behavior history information from application and video information in smartphone. Since many web pages are supposed to be published, deleting information on a specific personal web page makes it safe to use information on browsing history of web pages for authentication. The screen shot images that can be acquired when a web page browsing operation is performed using a standard browser on a smartphone have the following features.

1. Since browsing web pages by a browser is a user's voluntary behavior, the browsing history is information specialized for individual users

2. Since web pages are basically published to a large number of unspecified Internet users, it is difficult to identify the corresponding smartphone user only from the screen shot image of the web page

3. Services that use personal accounts tend to use dedicated applications, so browsers are more likely to perform simple browsing operations that do not link to personal information[13]

Since there are a large number of browsable web pages on the Internet, it is unlikely that the candidate images for authentication will be exhausted. Further, by using the URL recorded in the browsing history of the browser, the user's image registration work can be made unnecessary. In addition, since contents such as texts and images on web pages that the user has voluntarily browsed are stored in the memory of terminal, they are also effective for recall at authentication.

3.2 Research tasks

In image based authentication using behavior history, the selection method of correct image and incorrect images based on the selected information become problems.

The correct image used at authentication needs to be able to be recalled correctly from the memory by the user. In Ref.[7] [8], it is easy for the user to recall the correct image because users needs to register the correct image by themselves. However, in the case of image based authentication where the user does not register the authentication images, the user sees the correct image only after being presented on the authentication screen. Therefore, it is necessary to select the correct image which clearly remains in the user's memory.

On the other hand, incorrect images need to be easily distinguishable from the correct image for the user. In Ref. [7] [8], an image group other than the one registered by the user as a correct image is used as an incorrect images. When the user intentionally registers the correct image, it is easy to distinguish between the correct image and the incorrect images. However, if the incorrect images presented on the authentication screen is similar to the correct image, the authentication success rate decreases and the authentication operation time becomes longer. Therefore, it is necessary to prepare incorrect images looks significantly different to the correct image for the user himself. Also there is also a need to make few difference between the correct image and incorrect images for others.

4 Proposed method

4.1 Selection of correct image

There is an implicit method of acquiring information on the interest of each web page, without burdening the user, using the browsing time of the web page [14]. According to this method, the longer the web page is displayed, the more likely the user is interested in the page and the user is watching closely. It can be inferred that the user can easily remember the image of the web page that was watching closely. However, it is difficult to determine whether the user really watches closely at the screen, and an image not remaining in the memory of the user may be selected as the image of the

web page at which the user watched closely. Therefore, it is assumed that the user watches closely at the web page displayed on the screen of the smartphone while the operation of touching the screen with the thumb when browsing the web page.

$$\text{Gaze continuation rate} = \frac{\text{Touching time of the screen(ms)}}{\text{Browsing time of the web page(ms)}}$$

From the web page browsed by the user before closing the browser application, the web page with the highest gaze continuation rate is considered to be the most gazed, and the screen shot of that page is taken as the correct image.

4.2 Selection of incorrect images

4.2.1 Conditions to select incorrect images

It is important that the incorrect images are distinguishable from the correct image for the user and difficult to distinguish from the correct image for others. When a screen shot of a web page with low gaze continuation rate is selected as an incorrect image from browsed web pages, depending on the contents of the web page, the user may confuse the correct image with the incorrect image. Therefore, in order to prevent confusion of memory and to distinguish clearly from the correct image, screen shot images of the “unvisited” web page is selected as incorrect images.

4.2.2 Selection method excluding same websites as the correct image

As a first method, an incorrect images is selected from web page group obtained from another search result for the search keyword of the web page selected as the correct image. By selecting only an unvisited web page in this web page group as incorrect images, confusion of the user's memory is prevented, and it becomes easy for the user to distinguish between the correct image and the incorrect images. Figure 2 shows the flow of the selection method. First, a certain number of web page URL groups are extracted from the search result using the search keyword at the time of web page search selected as the correct image. The search result at this time may include the web page itself selected as the correct image, the viewed web pages, and web pages on the same domain name as those web pages. If these web page images are presented simultaneously at the time of image based authentication, the correct image may be confused with the incorrect images. Therefore, among the web page URL group obtained from the search result, URLs including the same domain name as the web page selected as the correct image and the browsed web page are excluded. And this selection method takes screen shots of the web pages from the URL of the remaining web pages and selects them as incorrect images.

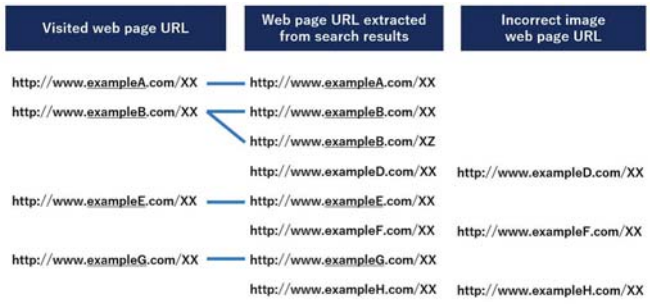


Figure 2: Selection method excluding the same website as the correct image

4.2.3 Selection method based on semantically distanced keywords

The second method is a method of extracting web pages that are semantically distant from web page that are correct images. By using the learning model using Word2Vec [15] as the search keyword of the web page URL selected as the correct image, it is possible to find a keyword that is semantically different from the search keyword of the correct image. Word2Vec takes a set of sentences as input and learns the vector representation of the word from other words that appear near the word. Since an incorrect images are selected using a search keywords different from the correct image, it is easy for the user to distinguish between the correct image and the incorrect images. Figure 3 shows the flow of the selection method. By calculating the cosine similarity with the search keyword of the web page selected as the correct image based on the language model learned by word2vec using the corpus in advance, a certain number of keyword groups with semantic distance apart are acquired. Screen shots of the web pages are randomly acquired one by one from the obtained search results of each keyword.

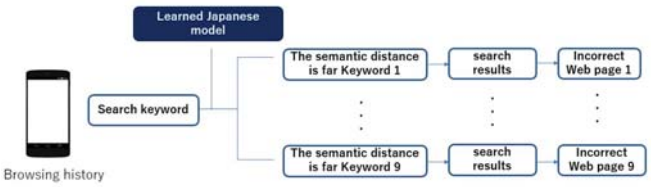


Figure 3: Selection method based on semantically distanced keywords

5 Evaluation experiment

The effectiveness of authentication and the validity of image selection are confirmed by evaluating the authentication success rate of the proposed system and the time taken for image selection at the time of authentication.

5.1 System configuration for experiment

In the evaluation experiment, a browser type data acquisition application for Android OS was developed and used as a system for the experiment. Table 1 shows the smartphones

used in the experiment, and Fig.4 shows the configuration of the experimental application. This application has a function of acquiring URLs, web page display time, screen contact time, gaze continuation rate, and search keywords. Each time the web page transitions, each data is sent to BrowserDB on the server. Google search engine was used for this search. In addition, the web page screenshots of the acquired URL are saved in the size of 320 x 568 pixels.

Table 1: Experimental device

Device name	VAIO Phone A VPA0511S
OS	Android 6.0.1
External dimensions	77.0 mm x 156.1 mm x 8.3 mm
Display size	5.5inch
Resolution	1080 x 1920

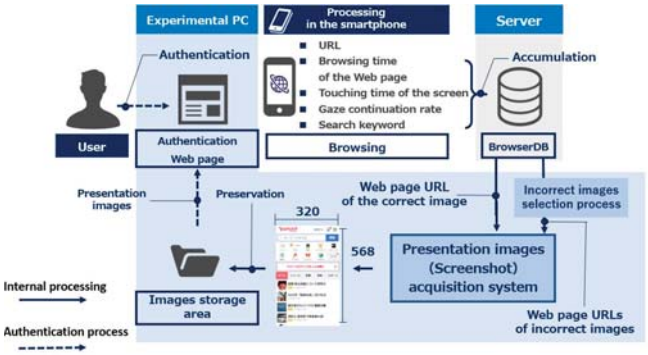


Figure 4: Experimental system

5.2 Experimental method

The subject browses the web page based on a search by a keyword designated in advance, and then performs an authentication operation of selecting a correct image from the authentication screen. On this authentication screen, 1 correct image selected by the correct image selection method and 9 incorrect images selected by the incorrect image selection method are displayed. In each trial of experience, the subject browses a web page for 10 minutes, takes 10 minutes break, and then performs an authentication operation. In order to measure the gaze continuation rate, and it is assumed that the thumb is always touching the screen when browsing web pages. In this method, the learned Japanese model[16] was used with Word2Vec. The search keywords to be specified are “カサブランカ (Casablanca)” and “端島 (Hashima)” in the trial using the incorrect image selection method excluding the same website as the correct image, and “土方歳三 (Toshizo Hijikata)” and “変身 (Henshin)” in the trial using the incorrect image selection method based on semantically distanced keywords between correct image. As keywords having a semantic distance from the search keyword at this time, the top 9 keywords with cosine similarity obtained by the learning model calculated by Word2Vec for each search keyword were used.

5.3 Result

In Fig.5 and Fig.6, the blue bar represents the result in the case of using Method A. The green bar represents the result in the case of using Method B. Figure 5 shows the average value of the authentication success rate in each method. When using Method A, the authentication success rate remains at 57.1%, whereas when using Method B, the authentication success rate has achieved 100.0%. Figure 6 shows the average image selection time in the authentication process for each method. The lines in the Fig.6 show the maximum and minimum values for each method. When using Method A, the average image selection time was 20.42 seconds, while when using Method B, it was significantly reduced to 9.04 seconds.

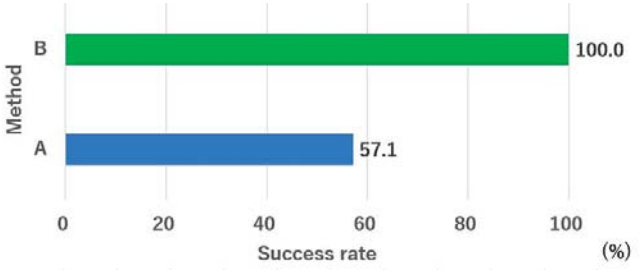


Figure 5: Authentication Success rate

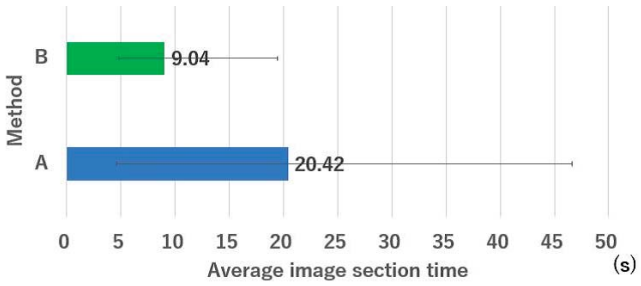


Figure 6: Average image section time

6 Discussion

Experimental results show that method B has a higher authentication success rate than method A. The cause of this is the content of the searched web pages, and the confusion of the user’s memory. Method A uses a search keyword when the correct image is browsed to select web pages that the user has not browsed as incorrect images. The content of the web pages of those incorrect images included the search keyword in the page, but the web pages of the content included in the search keyword were selected. For example, “カサブランカ (Casablanca)” is a web page such as an area or movie containing the keyword. For this reason, there are web pages that have similar searches but have similar contents, and I think that the user is confused. On the other hand, Method B uses the search keyword when viewing the correct image to acquire incorrect images using different keywords having a semantic distance. As a result, Web pages with content that was independent of search keywords were selected. For example, if it is “変身 (Henshin)”, these are web pages that contain

keywords such as “ロケット団’(Team Rocket)” and “カービィ(Kirby)”. For this reason, the user can easily distinguish the keyword from the content, and think that the correct image, which is the browsed web page, can be determined. This user’s memory confusion is considered to have an influence on the authentication speed, and it is considered that the authentication speed of Method B is faster than Method A, which easily causes memory confusion. Therefore, in this method, it can be said that by using an incorrect image selection method based on the semantic distance to the search keyword, it is possible to reduce user’s memory confusion and achieve a high authentication success rate.

7 Conclusion

In this paper, we proposed a personal authentication method that combines image based authentication, which is advantageous for reducing memory load and operation load, and behavior history authentication using personal data other than living body. There are problems are the selection of the correct image, and the selection of the incorrect images. The correct image used the screen shot image of the web page with the highest gaze continuation rate of the user. In selecting incorrect images, we proposed method A: selection method of incorrect images excluding same web page and method B: incorrect images selection method based on semantic distance to search keywords.As evaluation experiments, the effectiveness of two methods in choosing incorrect images was investigated. We performed experiments to select correct images from 1 correct image and 9 incorrect images by each method. In method A, the authentication success rate is 57.1 %, and the average image selection time is 20.42 seconds, but method B has 100.0% authentication success rate.The average image selection time was a result of 9.04 seconds. Therefore, it turned out that incorrect images selection method based on semantic distance to search keywords is effective.

As future tasks, we will investigate attack resistance in the Educated Guessing Attack, and discuss countermeasures for the vulnerable part.

REFERENCES

[1] splashdata, “The Top 50 Worst Passwords of 2018,” <https://www.teamsid.com/100-worst-passwords-top-50/> [Accessed June 6, 2019].

[2] statista, “Penetration of smartphones with fingerprint sensors worldwide from 2014 to 2018 ”, <https://www.statista.com/statistics/522058/global-smartphone-fingerprint-penetration/> [Accessed June 6, 2019].

[3] K. Cao A.K. Jain, “Hacking Mobole Phones Using 2D Printed Fingerprints,” MSU Technical Report, MSU-CSE-16-2, 2016.

[4] P. Bontrager, A. Roy, J. Togelius, N. D. Memon, and A. Ross, “DeepMasterPrints: Generating MasterPrints for Dictionary Attacks via Latent Variable Evolution,” Proceedings of the IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS2018), 2018.

[5] T. Takahashi, S. Kitagami, K. Miyashiro, E. T. Harada, and S. Suto, “Picture-authentication system(1):How users select registration pictures?,” Proceedings of the IEICE technical report. LOIS, Life intelligence and office information systems, Vol.112, No.35, pp.1-8, 2012(*in Japanese*).

[6] R. Niimi, A. Ueda, and K. Yokozawa, “Object perception, Series Integrated Perception,” Vol.2, pp.69-70, Keiso Shobo (2016). (*in Japanese*)

[7] R. Dhamija and A. Perrig, “Déjà Vu: A User Study Using Images for Authentication,” Proceedings of the 9th conference on USENIX Security Symposium(SSYM’00), pp.45-48, 2000.

[8] T. Takada and H. Koike, “Awase-E: the Method Enables an Image-based Authentication to be More Secure and Familiar for Users with Providing Image Registration and User Notification,” IPSJ Journal, Vol.44, No.8, pp.2002-2012, 2003(*in Japanese*).

[9] T. Ajioka, T. Umezawa, and N. Osawa, “Authentication Method for Mobile Device using Wrist Acceleration during Unlick Operation,” IPSJ SIG Technical Reports, Vol. 2016-MBL-81, No.22, pp.1-6,2016(*in Japanese*).

[10] K. Yamada, K. Notomi, and K. Saito, “Personal Authentication Method using Behavioral Feature Amounts of Smartphone Handling,” Proceedings of Journal of Biomedical Fuzzy Systems Association, Vol. 16, No.1, pp.41-48,2014(*in Japanese*).

[11] M. Nishigaki, and M. Koike, “A User Authentication Based On Personal History : A User Authentication System Using E-mail History,” IPSJ Journal, Vol.47, No.3, pp.945-956, 2006(*in Japanese*).

[12] N. Nguyen, and S. Sigg, “PassFrame: Generating image-based passwords from egocentric videos”, Proceedings of the 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), 2017.

[13] nielsen, ”Smartphone users who use the app and browser properly-Released “Digital Trends 2016 “ which summarized the trend of nielsen 2016-””, <https://www.netratings.co.jp/news_release/2017/03Newsrelease20170309.html/> [Accessed June 9, 2019](*in Japanese*).

[14] Y. Hijikata, “Exploiting Customer’s Preference - Leading Edge of User Profiling Technique- : Techniques of Preference Extraction for Information Recommendation,” IPSJ Magazine, Vol.48, No.9, pp.957-965, 2007(*in Japanese*).

[15] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J.Dean, “Distributed Representations of Words and Phrases andtheir Compositionality”, Neural Information Processing Systems2013, pp.3111–3119, 2013.

[16] AIAL, “Publish the learned Japanese model of word2vec,” <http://aial.shiroyagi.co.jp/2017/02/japanese-word2vec-model-builder/>(*in Japanese*) [Accessed October 29 2019].

On Theorem-Proving Trust Safety Properties – A Case Study

Toshinori Fukunaga[1], Hideki Goromaru[1], Tadanori Mizuno[2], Kazuhiko Ohkubo[1] and Yoshinobu Kawabe[2]

[1]NTT Secure Platform Laboratories, NTT Corporation
3-9-11 Midori-cho, Musashino-shi, Tokyo 180-0012, Japan
E-mail : { toshinori.fukunaga.vf, hideki.goroumaru.mx }@hco.ntt.co.jp,
ohkubo.kazuhiko@lab.ntt.co.jp
[2]Department of Information Science, Aichi Institute of Technology
Yachigusa 1247, Yakusa-cho, Toyota, Aichi 470-0392, Japan
E-mail : { tmizuno, kawabe }@aitech.ac.jp

Abstract - Information exchanged in social media may not be trustable. Some information may be true, but other information may be false. It is important to analyze the trust of information and the trust of information sources. Especially, in this paper we are interested in analyzing how the trust value changes as time passes. We discuss how to analyze the transitions of two-dimensional trust values. Specifically, we formalize time-related trust safety properties such as "The transition of trust value does not reach the region of distrust," and "If the trust value moves to the region of trust, then the trust value never goes to the region of distrust," in first-order logic, and we verify the correctness with a theorem-proving tool. A naive approach for the verification is to prove the first-order logic formulae directly, but this is not efficient. Hence, we apply a proof method for trace inclusion, which has been developed in the theory of distributed algorithms, to a verification of trust safety property. We conduct a case study with a simple example and we demonstrate a semi-automatic theorem-proving for the trust safety property.

Keywords: On-Line Trust, I/O-automaton, Safety Properties, Theorem-Proving

1 Introduction

Safety/relief information is actively exchanged via social media in recent large-scale disasters. Such information is not always trustable, and the correctness of information may change as time passes; i.e. even if a message "A person is seriously injured but currently alive," is true at the present time, it may become false one hour later. Thus, it is important to evaluate the trust of messages and the trust of information sources.

A (one-dimensional) trust value was introduced by Marsh and Dibben [1], and they classified the trust notions into trust, distrust, untrust and mistrust. This classification was based on the one-dimensional trust values, where the point of total trust and the point of total distrust are at the extremities. However, Lewicki indicated that trust and distrust should be treated as independent dimensions [2]. Moreover, the notion of trust is a concept closely related to human's impressions. Thus, "contradictions/confusions" and "ignorance" should be considered in evaluating the trust values. From this viewpoint, we introduced a two-dimensional trust representation with a pair of trust value and distrust value [3][4]. We employed a theory for impression formation (Oda's FCR method [5][6][7]) as a basis for the trust representation, and the correspondence between our trust representation and Marsh and Dibben's conventional representation was explored.

The two-dimensional trust value of [3][4] represents a trust state at a certain moment. However, it is necessary to handle the changing nature of trust values to analyze trust-related propertie. Thus, a property of ever-changing trust value is modeled in this paper. Specifically, a safety property is defined with the transition sequences of trust values. Furthermore, we conduct an efficient computer-assisted proof for the trust safety property based on the results in I/O-automaton theory [8][9].

2 Two-Dimensional Trust Representation

The trust classification by Marsh and Dibben is as follows, where a trust value ranges over $[-1, 1)$:

- *Trust*: is a state where a trust value of a trustee is more than a threshold value;
- *Distrust*: is a state where the trust value is negative;
- *Untrust*: is a state where the trustee's trust value is positive but not enough to cooperate; and
- *Mistrust*: is a state in which the initial trust has been betrayed; more precisely, the notion of mistrust can be considered as "Either a former trust destroyed, or former distrust healed," since the trustee may not have had bad intentions and it is not always "betrayed".

Suppose that you received a message, and you calculated its trust value. If the trust value is 0.9 and the cooperation threshold is 0.85, then from the definition of the trust notion, the message should be trusted. However, can we say that there is no distrust on this message? The maximum of the trust value is 1, hence we can see that there is a deficit of 0.1 points on the trust value. In this sense, the message might not be trusted enough. In [3][4], we considered that this was due to the limitation on the expressive power of one-dimensional trust representation, and we introduced a two-dimensional trust value to be an element of $Trust \times DisTrust$, where $Trust$ and $DisTrust$ are respectively degrees of trust and distrust, and

The fourth author is currently at KYOWA EXEO CORPORATION.

we have $Trust = DisTrust = \{\, v \,|\, 0 \leq v \leq 1\}$. Following the manner in the FCR method, a two-dimensional trust value is also called an observation.

An observation around $(1, 0)$ has a high trust value and a low distrust value. Thus, we can see that the observation represents a state of "trust". Similarly, observations around $(0, 1)$ are the states of "distrust" since they have a low trust value and a high distrust value. For any observation (t, d) on the line between $(1, 0)$ and $(0, 1)$, we can see that (t, d) is ideal in the sense that the trust value and the distrust value satisfy the consistency condition $t + d = 1$. We consider that Marsh and Dibben's trust values are on this line. That is, the conventional trust value is defined with a limitation with regard to the consistency condition. Finally, the observation which corresponds to the conventional trust value of 0 is $(0.5, 0.5)$.

We defined a classification of trust for two-dimensional trust values with a transformation:

$$\left[\begin{pmatrix} \cos\frac{\pi}{4} & -\sin\frac{\pi}{4} \\ \sin\frac{\pi}{4} & \cos\frac{\pi}{4} \end{pmatrix} \left\{ \begin{pmatrix} t \\ d \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\} + \begin{pmatrix} \frac{\sqrt{2}}{2} \\ 0 \end{pmatrix}\right] \times \frac{1}{\frac{\sqrt{2}}{2}} = \begin{pmatrix} t-d \\ t+d-1 \end{pmatrix}$$

where the resulting point $(t - d,\ t + d - 1)$ is called (i, c).

First, we consider the first element $i = t - d$ of (i, c). This is a value with $-1 \leq i \leq 1$, and the value corresponds to the conventional trust value of Marsh and Dibben. In fact, the value i indicates that the result of subtracting the degree of trust by the degree of distrust is actually the net trust value, and we see that this matches the intuition.

The value of c is the degree of contradiction-irrelevance $C(t, d)$ defined in the theory of FCR method. The degree of contradiction represents how much it deviates from the consistent condition ($t + d = 1$), and its value is between -1 and 1. The degree of contradiction is close to 1 if we deal with a trustee of a contradictory evaluation; for example, "I trust him but at the same time I feel some distrust on his behaviour". Also, if the degree is around -1, then a truster is ignorant on a trustee; that is, this is a situation like "I do not care for him at all." If we have $t + d = 1$, then the degree of contradiction is 0.

3 Transitions of Observations

Mistrust is a property with regard to a misplaced trust, and this is related to a change of trust values over time. In order to build a trust relationship among victims and volunteers in a large-scale disaster [10][11][12], it is important to analyze a chronological change of trust values. In this section, we deal with the observations in the previous section as states, and we formalize time-related trust properties with state machines.

I/O-automaton [8][9] is a mathematical model for distributed algorithms. In the theory of I/O-automata, a system is regarded as a collection of state machines which interacts with each other; the interactions are formalized with events. Some of events are observable from the outside of the system, and some events cannot be observed. A sequence of observable events from the initial state is called a trace, and a set of traces characterizes the system. I/O-automata have multiple (possibly infinite) traces, and the properties of the automata are characterized by the set of traces.

Safety and liveness properties are well-known properties of distributed algorithms and they are defined with traces. A safety property guarantees that there is no occurrence of (specified) bad event. For example, if a computer program has no "division by 0" errors, then we regard that the program satisfies a safety property. On the other hand, a liveness property represents that finally some good behaviour will happen. For example, if a computer program always terminates, we can see that the program has a liveness property. If a communication system can reach an initial state from any state of the system, then we can see that the communication system satisfies another liveness property.

If we regard observations as states, the property "The trustee never goes to the region of distrust," is regarded as a safety property on trust transitions. Also, "A trustee will finally reach the region of trust," can be considered an liveness property on trust. In the following, let CT be a cooperation threshold with $0 < CT \leq 1$. We define the trust region $T(CT)$, the distrust region D, and the untrust region $U(CT)$ with:

$$\begin{cases} T(CT) &= \{\, (t,d) \,|\, t \in Trust \wedge d \in DisTrust \\ & \quad \wedge t - d \geq CT \,\}, \\ D &= \{\, (t,d) \,|\, t \in Trust \wedge d \in DisTrust \\ & \quad \wedge t < d \,\}, \text{ and} \\ U(CT) &= Trust \times DisTrust \setminus (T(CT) \cup D) \end{cases}$$

and we formalize trust safety properties.

Formally, automaton X has a set of actions $sig(X)$, a set of states $states(X)$, a set of initial states $start(X) \subset states(X)$ and a set of transitions $trans(X) \subset states(X) \times sig(X) \times states(X)$. Transition $(s, a, s') \in trans(X)$ is written as $s \xrightarrow{a}_X s'$. In this paper, a state is a tuple of values. Each element of the tuple has a corresponding distinct variable name. The name of a variable is used as an access function to the value. This kind of modeling is standard in I/O-automaton theory and its extensions such as [13]. In this paper, we use variables tr and dis for trust value and distrust value, respectively. The degrees of trust and distrust in state $s \in states(X)$ are referred as $s.tr$ and $s.dis$, respectively.

For any state $s \in states(X)$, a property "If s is not in the distrust region then the next state of s is not in the distrust region," is defined with:

$$\begin{aligned} &stepTrustSafe(s) \Longleftrightarrow \\ &\quad (s.tr,\ s.dis) \notin D \\ &\qquad \Longrightarrow\ \forall a \in sig(X)\, \forall s' \in states(X) \\ &\qquad\qquad [\, s \xrightarrow{a}_X s' \ \Longrightarrow\ (s'.tr,\ s'.dis) \notin D \,]. \end{aligned}$$

Hence, if we prove

$$\begin{aligned} &\forall s \in start(X) [\, (s.tr, s'.tr) \notin D \,] \\ \wedge\ &\forall s \in state(X) [\, stepTrustSafe(s) \,] \end{aligned} \tag{1}$$

then we have "The system X never reaches the distrust region." This formula consists of two conditions. The first condition represents that an initial state is not in the distrust region. The second condition means that every state s should

satisfy $stepTrustSafe(s)$; that is, for any transition from s the system X never goes to the distrust region. If we use the predicate $reachable(s, s')$ for the reachability from state s to state s', the second condition can be:

$$\forall s_{init} \in start(X), \forall s \in state(X) \\ [\, reachable(s_{init}, s) \implies stepTrustSafe(s) \,]$$

In this case, we consider the safety property only for reachable states. In any cases, this is to prove a trust safety property by induction on the length of execution sequences.

With the predicate $reachable$, another safety property "If a user exits the region of distrust, then the user never goes back to the distrust region," is formalized with:

$$\forall s, s' \in states(X) \\ [(reachable(s, s') \wedge (s'.tr, s'.dis) \notin D) \\ \implies \forall s'' \in states(X) \\ [reachable(s', s'') \\ \implies (s''.tr, s''.dis) \notin D]\,].$$

4 Proving Trust Safety Property

Let X be an automaton which specifies a communication system. To prove a trust safety property of X, it suffices to prove the condition (1) in the previous section with a theorem-proving tool directly. However, this approach is not efficient.

In this section, we apply an efficient proof method for trace inclusion of I/O-automaton to the verification of trust safety properties. Specifically, we describe two automata. The first automaton is what we can easily check that the automaton satisfies a trust safety property. The second automaton is a specification of the target communication system. If we can prove the trace inclusion between the two automata, the trust safety property of the first automaton leads to the trust safety property of the second automaton.

4.1 Preliminary Definition

First, we introduce the definition for sort `VL` with the Larch language. Specifically, the definition is:

```
VL: trait

  introduces
    vl0, vl1, otherValues: -> VL
    -__, abs: VL           -> VL
    __+__, __-__: VL, VL   -> VL
    __<__, __<=__,
    __>__, __>=__: VL, VL  -> Bool
    CT:                    -> VL
    D: VL, VL              -> Bool
    T: VL, VL, VL          -> Bool
    U: VL, VL, VL          -> Bool

  asserts with x, y, z, t, d, ct: VL
    x + vl0 = x;
    vl0 + x = x;
    x - vl0 = x;
    vl0 - x = -x;
    -(-x) = x;
    (x >= vl0) => abs(x)  = x;
    (x < vl0) => abs(x)  = -x;
    abs(-x) = abs(x);
    (-vl1) <= x;
    x <= vl1;
    (x >= y) <=> ((x = y) \/ (x > y));
    (x <= y) <=> ((x = y) \/ (x < y));
    (x > y) <=> ~(x <= y);
```

```
uses testerSafetyDT

automaton testerSafety
  signature
    internal move(ev:Event, pt: VL,
                  pd: VL, dt: VL, dd: VL)
    output inDistr(t:VL, d:VL)
    output notInDistr(t:VL, d:VL)

  states
    tr: VL := vl0,
    dis: VL := vl0,
    stateOfAgent: agtState := InitState

  transitions
    internal move(ev, pt, pd, dt, dd)
      pre    pt = tr
          /\ pd = dis
          /\ (   vl0 <= (pt + dt)
              /\ (pt + dt) <= vl1)
          /\ (   vl0 <= (pd + dd)
              /\ (pd + dd) <= vl1)
          /\ condition(stateOfAgent, ev,
                       pt, pd, dt, dd)
      eff tr  := tr + dt;
          dis := dis + dd;
          stateOfAgent
              := change(stateOfAgent, ev)

    output inDistr(t, d)
      pre tr < dis /\ t = tr /\ d = dis
      eff tr := tr

    output notInDistr(t, d)
      pre ~(tr < dis) /\ t = tr /\ d = dis
      eff tr := tr
```

Figure 1: `testerSafety`: An Abstract System

```
(x < y) <=> ~(x >= y);
CT > vl0;
CT < vl1;
D(t, d) <=> ((t-d) < vl0);
T(t, d, ct) <=> ((t-d) >= ct);
U(t, d, ct)
    <=> (~D(t, d) /\ ~(T(t, d, ct)))
```

and this kind of description is called a trait. The trait `VL` is for the set of real numbers whose domain is $[-1, 1]$. In this trait, several constants and operators are introduced; for example, constants `vl0` and `vl1` are respectively for 0 and 1 in sort `VL`. `CT` is a term for the cooperation threshold. Addition `+`, subtraction `-`, an unary operator for negative numbers `-`, comparison operators `<`, `>`, `<=` and `>=`, and the function for absolute value `abs` are defined as usual.

Predicates `D(t, d)`, `T(t, d, ct)` and `U(t, d, ct)` are true if observation `(t, d)` is in the distrust region, the trust region and the untrust region, respectively; note that `ct` is a parameter for the cooperation threshold.

4.2 Specifying And Proving Safety Property

We introduce Fig. 1's I/O-automaton `testerSafety` to define a trust safety property. The automaton has three actions:

- `move(ev, pt, pd, dt, dd)`: enabled if event `ev` occurs and the two-dimensional trust value changes from `(pt, pd)` to `(pt+dt, pd+dd)`;

```
testerSafetyDT: trait

includes VL

introduces
  condition: agtState, Event,
             VL, VL, VL, VL -> Bool,
  InitState: -> agtState,
  AnotherState: -> agtState,
  change: agtState, Event -> agtState

asserts with st: agtState, ev: Event,
            pt, pd, dt, dd: VL
  condition(st, ev, pt, pd, dt, dd)
       <=> ~D(pt, pd) /\ ~D(pt+dt, pd+dd)
```

Figure 2: Datatype for `testerSafety`

- `inDistr(t, d)`: enabled if trust value `(t, d)` is in the distrust region; and
- `notInDistr(t, d)`: enabled if trust value `(t, d)` is not in the distrust region.

Note that actions `inDistr` and `notInDistr` are special actions for analyzing trust transitions. If action `inDistr` does not appear on any trace, the automaton will not go to the distrusted region.

The transition of trust values is determined only by action `move`, and the action is enabled if predicate `condition` in the `pre`-part is true. The predication `condition` is introduced in Fig. 2's `testerSafetyDT`, and it is defined as

```
condition(st, ev, pt, pd, dt, dd)
    <=> ~D(pt, pd) /\ ~D(pt+dt, pd+dd)
```

in this study. This condition means that "The observation `(pt, pd)` of the current state and the observation `(pt + dt, pd + dd)` of the next state are neither in the distrust region." It is important for verifiers that the correctness of automaton `testerSafety` can be checked easily. Actually, `testerSafety` is small and simple. Moreover, with the `condition` predicate above, we can easily see that there is no occurrence of `inDistr` in `testerSafety`'s traces; this leads to the correctness of `testerSafety`.

An example of event sequence from `testerSafety` is:

```
notInDistr(vl0, vl0).
  move(get_mes(usrA, "hello"),
               vl0, vl0, vl0.3, vl0.2).
    notInDistr(vl0.3, vl0.2).
```

In the initial state, both of the trust and distrust degrees are 0; that is, we have `(tr, dis) = (vl0, vl0)`. In this case, `tr < dis` holds, and this allows an occurrence of event `notInDistr(vl0, vl0)`. Then, after the occurrence of event `get_mes`, the pair of trust and distrust degrees becomes `(vl0.3, vl0.2)`, where constants `vl0.3` and `vl0.2` in sort `VL` represent 0.3 and 0.2 respectively. In the resulting state, the distrust degree still does not exceed the trust degree, thus event `notInDistr(vl0.3, vl0.2)` can occur.

In the specification of Fig. 1, we have variables `dt` and `dd` in action `move`. They are variables of sort `VL`, and the definitions in trait `VL` restrict their values to be in $[-1, 1]$. Additionally, `dt` and `dd` should be the values which satisfy both of values `pt + dt` and `pd + dd` are in $[0, 1]$. This condition is written in the precondition part of `move`.

Automaton `testerSafety` can be translated into first-order logic formulae, and a safety property can be proven with Larch Prover (LP) [14]. In the following we prove the safety property that "For any state `s` of `testerSafety`, if `s` is reachable then the two-dimensional trust value at `s` is not in the distrust region." This is formally described as:

```
(\A st:States[testerSafety]
    (reachableAbst(st)
     => ~D(st.tr, st.dis))).
```

Proving this formula is equivalent to proving two conditions

```
  (\A st:States[testerSafety]
      (start(st) => ~D(st.tr, st.dis)))
```

and

```
(\A st:States[testerSafety]
(\A at:Actions[testerSafety]
    (reachableAbst(st)
     => ((   enabled(st, at)
          /\ ~D(st.tr, st.dis))
         => ~D(effect(st, at).tr,
               effect(st, at).dis))))).
```

This is to prove a trust safety property by induction on the length of execution sequences; the first condition represents a base case, and the second condition is an induction step. From this, we have `~D(st.tr, st.dis)` for any reachable state `st` of `testerSafety`. This means `inDistr` is not enabled at `st`. Therefore, we obtain the following lemma.

Lemma 1 *Every trace of I/O-automaton* `testerSafety` *does not have any occurrence of action* `inDistr`. □

In this sense, `testerSafety` never goes to the region of distrust; this leads to a trust safety property.

4.3 Specifying Communication System And Safety Proof by Trace Inclusion

4.3.1 Specification of Communication System

We consider a specification `bbdSystem` of a communication system; see Fig. 3. This communication system sends a message to an online bulletin board after evaluating the user's trust value. The system receives a message from a user by action `get_mes` and evaluates the trust value of the message with actions `discard_mes` and `approve_mes`. If the trust value in the next state reaches the distrust region when writing the message to the bulletin board, the message is not sent and discarded by `discard_mes`. Otherwise, the message is written to the bulletin board by actions `approve_mes` and `say`. Actions `inDistrC` and `notInDistrC` are special actions to discuss the trust values, and they correspond to actions `inDistr` and `notInDistr` of `testerSafety`.

Automaton `bbdSystem` uses a datatype defined in the trait `bbdSystemDT` shown in Fig. 4. This trait employs `Sequence(MES)`, which defines a message queue.

```
uses bbdSystemDT

automaton bbdSystem
  signature
    input get_mes(i:ID, m:MES)
    internal discard_mes(i:ID, m:MES)
    internal approve_mes(i:ID, m:MES)
    output say(i:ID, m:MES)
    output inDistrC(t:VL, d:VL)
    output notInDistrC(t:VL, d:VL)

  states
    tr: VL := vl0,
    dis: VL := vl0,
    flg: Bool := false,
    mesQ: Seq[MES] := empty

  transitions
    input get_mes(i, m)
      eff mesQ := mesQ ||
                  (packet(i, m) -| empty)

    internal discard_mes(i, m)
      pre    ~flg
          /\ mesQ ~= empty
          /\ packet(i, m) = head(mesQ)
          /\ ((tr + evalTr(tr, m))
              -(dis + evalDis(dis, m)))
             < vl0
      eff mesQ := tail(mesQ)

    internal approve_mes(i, m)
      pre    ~flg
          /\ mesQ ~= empty
          /\ packet(i, m) = head(mesQ)
          /\ ((tr + evalTr(tr, m))
              -(dis + evalDis(dis, m)))
             >= vl0
      eff flg := true

    output say(i, m)
      pre flg /\ mesQ ~= empty
              /\ packet(i, m) = head(mesQ)
      eff tr := tr + evalTr(tr, m);
          dis := dis + evalDis(dis, m);
          mesQ := tail(mesQ);
          flg := false

    output inDistrC(t, d)
      pre tr < dis /\ t = tr /\ d = dis
      eff tr := tr

    output notInDistrC(t, d)
      pre ~(tr < dis) /\ t = tr /\ d = dis
      eff tr := tr
```

Figure 3: `bbdSystem`: A Concrete System

We introduce functions `evalTr` and `evalDis` for calculating the degree of trust and the degree of distrust, respectively, though the concrete definitions of these functions are not given in this paper; we only give the constraints on these functions. For function `evalTr`, we employ a constraint

```
(\A tr:VL
    ((vl0 <= tr /\ tr <= vl1)
     => (\A m:MES
            (   vl0 <=
                (tr + evalTr(tr, m))
             /\ (tr + evalTr(tr, m))
                <= vl1))))
```

for the degree of trust to be in $[0, 1]$. For `evalDis`, a similar condition is introduced.

4.3.2 Safety Proof by Trace Inclusion

We replace `bbdSystem`'s observable actions `get_mes` and `say` with internal actions of the same name; the resulting

```
bbdSystemDT: trait

includes VL, Sequence(MES)

introduces
  packet:  ID, MES -> MES
  evalTr:  VL, MES -> VL
  evalDis: VL, MES -> VL
  getMesFromPacket:  MES -> MES

asserts with tr, dis: VL, i:ID, m: MES
  (\A tr:VL
      ((vl0 <= tr /\ tr <= vl1)
       => (\A m:MES
              (   vl0 <=
                  (tr + evalTr(tr, m))
               /\ (tr + evalTr(tr, m))
                  <= vl1))));
  (\A dis:VL
      ((vl0 <= dis /\ dis <= vl1)
       => (\A m:MES
              (   vl0 <=
                  (dis + evalDis(dis, m))
               /\ (dis + evalDis(tr, m))
                  <= vl1))));
  getMesFromPacket(packet(i, m)) = m;
```

Figure 4: Datatype for `bbdSystem`

automaton is called $\texttt{bbdSystem}\backslash\{\texttt{get\_mes}, \texttt{say}\}$. If we prove the existence of a binary relation called a forward simulation from automaton $\texttt{bbdSystem}\backslash\{\texttt{get\_mes}, \texttt{say}\}$ to automaton `testerSafety`, then we have a trace inclusion

$$traces(\texttt{bbdSystem}\backslash\{\texttt{get\_mes}, \texttt{say}\}) \subseteq traces(\texttt{testerSafety})$$

from Theorem 3.10 of [9]. From Lemma 1, every trace in $traces(\texttt{testerSafety})$ does not have any occurrence of action `inDistr`. Hence, if we prove a trace inclusion, every trace in $traces(\texttt{bbdSystem}\backslash\{\texttt{get\_mes}, \texttt{say}\})$ does not contain `inDistr`; this leads to the absence of `inDistr` in $traces(\texttt{bbdSystem})$.

A candidate binary relation for a forward simulation is defined as follows:

```
 fs(sb, st) <=> (   (sb.tr = st.tr)
                 /\ (sb.dis = st.dis)
                 /\ (sb.flg => X) )

where X =
 ( ((sb.tr
     + evalTr(sb.tr,
         getMesFromPacket(head(sb.mesQ))))
   - (sb.dis
      + evalDis(sb.dis,
         getMesFromPacket(head(sb.mesQ)))))
   >= vl0)
```

To prove that binary relation `fs` is a forward simulation, we should prove the initial state condition

```
 (\A sb:States[bbdSystem]
     (start(sb)
      => (\E st:States[testerSafety]
             (start(st) /\ fs(sb, st)))))
```

and step correspondence condition

```
(\A sb:States[bbdSystem]
(\A sb':States[bbdSystem]
(\A st:States[testerSafety]
(\A ab:Actions[bbdSystem]
    (reachableAbst(st)
     => ((fs(sb, st) /\ step(sb, ab, sb'))
         => (\E st':States[testerSafety]
                (   steps(st, ab, st')
                 /\ fs(sb', st')))))))))
```

with a theorem proving tool. This leads to the following lemma.

Lemma 2 *Binary relation* `fs(sb, st)` *is a forward simulation from automaton* `bbdSystem\{get_mes, say}` *to automaton* `testerSafety`. □

Summarizing, we have the following result.

Theorem 1 *Every trace of I/O-automaton* `bbdSystem` *does not have an occurrence of action* `inDistr`.

Proof: Proven by Lemmata 1 and 2. □

Consequently, a trust safety property has been shown for automaton `bbdSystem`.

5 Conclusion

In this paper, we discussed how to analyze transitions of two-dimensional trust values. Specifically, we employed a theory of distributed algorithms to formalize trust safety properties such as "Any transition does not lead to the distrust region," or "After reaching some region other than the distrust region, the system never goes to the distrust region."

We cannot say that it efficient to theorem-prove the condition for trust safety property of each automaton. In this paper, we employed I/O-automata to represent a trust safety property, and applied a proof method for trace inclusion. This enables us to verify trust safety properties efficiently. Furthermore, with simple examples, we empirically demonstrated that it is possible to verify trust safety properties with automatic theorem provers.

REFERENCES

[1] S. Marsh and M. R. Dibben, "Trust, untrust, distrust and mistrust – an exploration of the dark(er) side," in *Proceedings of the Third International Conference on Trust Management*, iTrust'05, (Berlin, Heidelberg), pp. 17–33, Springer-Verlag, 2005.

[2] R. J. Lewicki, D. J. B. McAllister, and R. J. Bies, "Trust and distrust: New relationships and realities," *Academy of Management Review*, vol. 23, pp. 438–458, 1998.

[3] K. Ohkubo, T. Oda, Y. Koizumi, T. Ohki, M. Nishigaki, T. Hasegawa, and Y. Kawabe, "Trust representation under confusion and ignorance," in *Proceedings of International Workshop on Informatics (IWIN 2018)*, pp. 191–198, 2018.

[4] Y. Kawabe, Y. Koizumi, T. Ohki, M. Nishigaki, T. Hasegawa, and T. Oda, "On trust confusional, trust ignorant, and trust transitions," in *Proceedings of IFIPTM 2019*, 2019.

[5] T. Oda, "Fundamental characterestics of fuzzy-set concurrent rating method," *Journal of Japan Association for Management Systems*, vol. 12, no. 1, pp. 23–32, 1995. In Japanese.

[6] T. Oda, "Fuzzy set theoretical approach for improving the rating scale method : Proposing and introducing the FCR-method and the IR-method as novel rating methods," *Japanese Psychological Review*, vol. 56, no. 1, pp. 67–83, 2013. In Japanese.

[7] T. Oda, "Measurement technique for ergonomics, section 3: Psychological measurements and analyses (3) measurements and analyses by kansei evaluation," *The Japanese Journal of Ergonomics*, vol. 51, no. 5, pp. 293–303, 2015. In Japanese.

[8] N. A. Lynch, *Distributed Algorithms*. Morgan Kaufmann Publishers, 1996.

[9] N. Lynch and F. Vaandrager, "Forward and backward simulations — part I: Untimed systems," *Information and Computation*, vol. 121, pp. 214–233, Sept. 1995.

[10] Y. Murayama, "Issues in disaster communications," *Journal of Information Processing*, vol. 22, no. 4, pp. 558–565, 2014.

[11] M. G. Busa, M. T. Musacchio, S. Finan, and C. Fennell, "Trust-building through social media communications in disaster management," in *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, (New York, NY, USA), pp. 1179–1184, ACM, 2015.

[12] F. Lemieux, "The impact of a natural disaster on altruistic behaviour and crime," *Disasters*, vol. 38, pp. 483–499, July 2014.

[13] D. Kaynar, N. Lynch, R. Segala, and F. Vaandrager, *The Theory of Timed I/O Automata, Second Edition*. Morgan & Claypool Publishers, 2nd ed., 2010.

[14] J. F. Soegaard-Andersen, S. J. Garland, J. V. Guttag, N. A. Lynch, and A. Pogosyants, "Computer-assisted simulation proofs," in *CAV '93*, vol. 697 of *Lecture Notes in Computer Science*, pp. 305–319, Springer-Verlag, 1993.

Keynote Speech 1:
Dr. Phil. Tim Ziemer
(Researcher at the University of Bremen)

Monophonic Sonification for Spatial Navigation

Tim Ziemer, Thomas Höring, Lukas Meirose and Holger Schultheis

* University of Bremen, Bremen, Germany
{ziemer, thoering, meirose, schulth}@uni-bremen.de

Abstract - Sonification is the communication of variable data by means of sound. It can be used as an alternative or augmentation of visualization. In this paper we describe and evaluate a sonification approach that communicates the location of a target in three dimensional space. This is achieved by thorough consideration of auditory perception in our sonification. An experiment with passive listeners reveals that the sonification is readily interpretable. Listeners are able to identify the direction and distance along the single dimensions. This renders the sonification suitable for interactive navigation tasks in three dimensional space.

Keywords: sonification, auditory display, digital signal processing, psychoacoustics, human-computer interaction

1 INTRODUCTION

Sonification is a systematic transformation of data to sound. An overview can be found in *The Sonification Handbook* [1]. Interactive sonification is the real-time sonification of data that can be manipulated by the user. In the case of navigation sonification the location of a target relative to the location of the user is sonified. Navigation sonification guides the user towards a desired target location. Several approaches towards navigation in one-[2], two-[3, 4], and three dimensional [5, 6] space exist. Some leverage spatial audio; the location of the target is represented by a virtual sound source whose location coincides with the target location. This can be achieved using headphones [7] or loudspeaker arrays [6]. An alternative is to use monophonic sound [2, 8]; different characteristics of the sound represent different directions. The magnitude of the characteristic represents the distance along that direction. Potential application areas include guidance of pedestrians [7], image guided surgery and neuronavigation [9, 10], surgical training [11], piloting [12], car parking assistance [13], and audible motor feedback for neuromotor rehabilitation [14, 15].

To date, virtually no navigation sonification method fulfills all requirements for precise and accurate navigation in three dimensional space, which are

1. Three Dimensions
2. Orthogonality
3. Clear Reference
4. High Resolution .

For three dimensional navigation, the three spatial dimensions need to be represented by means of at least three characteristics of sound. Assigning spatial input data to a certain audio output parameter is referred to as mapping. Many navigation sonification approaches are restricted to one or two dimensions [2, 3, 4]. One reason for that may be the fact that it is difficult to create a sound with orthogonal characteristics. Here, orthogonality does not refer to independent audio parameters but to independent listening impressions, as described in psychoacoustic literature [16]. Orthogonality means that each orthogonal spatial dimension has to be represented by a sound attribute, which is independent of the attributes that represent other dimensions. Physically orthogonal attributes of sound, like frequency and amplitude, interfere perceptually. Both can affect the perceived pitch and loudness. Hence, sonification designers use psychoacoustic signal processing [5, 17] to map dimensions to orthogonal perceptual sound attributes, like pitch/loudness/brightness [15] or the three perceived spatial dimensions [6]. The first approach has the disadvantage that there is no natural zero for pitch, loudness, and brightness. Therefore, the target pitch, loudness, and brightness have to be pre-defined. Then, either the user learns and memorizes them to navigate, or the navigation sonification is interrupted to play a reference sound, i.e., the target sound. This way the user can compare the two to derive the direction and distance of the target. In the first case, there is no clear reference, because our memory for absolute pitch, loudness and brightness is weak. In the latter case, the temporal resolution is degraded because the sonification is interrupted by the target sound. When leveraging spatial audio for navigation as in [7, 6], the resolution is relatively low along some dimensions: even though we may be able to distinguish about 50 azimuth angles, the resolution along the altitude dimension and the distance is much lower [18]. That is the reason why [6] added monaural cues in addition to spatial audio for three dimensional navigation.

Recently, we presented a monophonic sonification approach which aims at fulfilling all the above-mentioned requirements. Details and demonstration videos can be found in [5]. The paper at hand reports initial results of a user study to evaluate the approach. First, we briefly describe the mapping principles between physical directions and distances and audible sound characteristics. Then, we describe the experiment setup, followed by a presentation of the results, a brief discussion and a conclusion.

2 SONIFICATION

The monophonic sonification for navigation in three dimensional space maps spatial directions to different characteristics of a sound. The distance along each direction is represented by the magnitude of each characteristic. Details of the implementation, including the psychoacoustic signal pro-

cessing and some demonstration videos, can be found in [5]. The sonification core resembles our two dimensional sonification [4, 3]. However, a third dimension has been added.

Figure 1 visualizes the sonification principle metaphorically. It illustrates the current location of a user at the origin of a coordinate system. The current location is represented by a red cursor. The graphic shows how the sonification sounds when a target lies at the respective location within the three dimensional space.

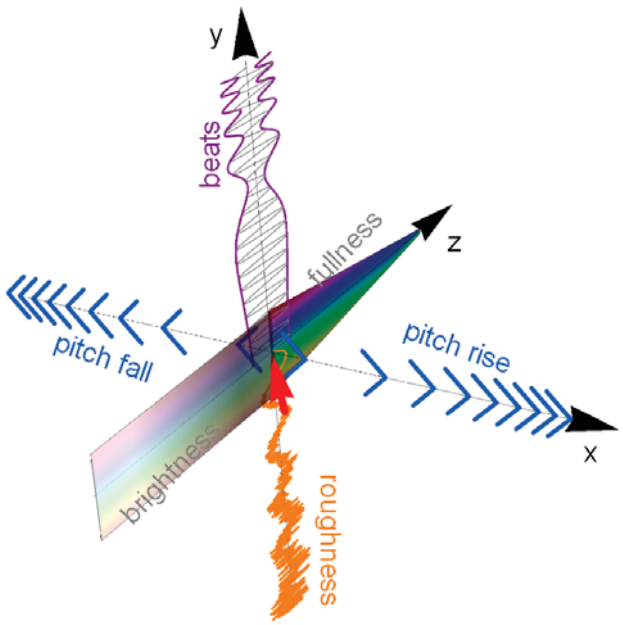


Figure 1: Sonification of a target location relative to the current location (red cursor). The x-dimension is related to pitch, the y-dimension to beats and roughness and the z-dimension to brightness and fullness.

The x-dimension is represented by pitch change [19]. When a target lies to the right, the pitch rises. The further to the right, the faster the pitch rises. When it lies to the left, the pitch falls accordingly. Only at the target x-location the pitch is steady. A so-called *Shepard tone* [20] is leveraged to achieve this. It can create the auditory illusion of an infinite pitch rise or fall. In contrast to pitch itself, pitch change has a natural zero, which is a steady pitch.

The y-dimension is divided in two: when the target lies above, the loudness of the sonification will alter frequently, referred to as *beats* [19]. The beat frequency communicates how far above the target lies. The faster the beats, the further does the target lie above. The sensation of beats results from slow amplitude modulations of traveling waves inside the cochlea, which can be created by slow (< 10 Hz) amplitude modulations of audio signals [21, ch. 4]. Just like pitch change, beats have a natural zero, which is a constant loudness. When the target lies below, the sonification will sound *rough* [22, 23]. Roughness is also referred to as *sensory dissonance* [19]. Roughness is a sensation that arises from fast ($15 < f_\mathrm{m} < 150$ Hz) amplitude fluctuations f_m of the traveling waves inside the cochlea. Such fluctuations can be created by means of frequency modulation with a constant modulation frequency of 50 Hz. All carrier frequencies are modulated by $f_\mathrm{m} = 50$ Hz, which creates sidebands equally distributed around the carrier frequencies. Even though roughness does have a natural zero, it is contextual: After hearing a very rough sound, even mild roughness may appear as perfectly smooth to a listener. After hearing a perfectly smooth sound, the same mild roughness may be obvious for the listener.

The functionality of these two dimensions has already been demonstrated in experiments with passive listeners [24, 25] and active users [3, 4]. Already passive users were able to attribute a sonified target to the correct region on a map with 16 regions. Here, the average hit-rate was 41%, while the chances to guess the correct field were 6.25%. In an interactive experiment users found no less than 92% of the targets.

The z-dimension is also divided in two: when the target lies in front, the sonification will sound less *full* [22]. The sensation of auditory fullness is related to its bandwidth. Fullness is also referred to as *volume* or *sonority* [22]. At the target the sonification sounds full. It contains 12 frequencies. The further away the target lies to the front, the lower the amplitude of the outermost frequencies becomes. A target far to the front contains three carrier frequencies only. When the target lies in the rear, the sound becomes *bright*. Brightness is also referred to as *sharpness* [16, 23] and is closely related to the spectral centroid; the further to the rear, the brighter the sound. Both fullness and brightness only exhibit a vague natural zero. This means the target fullness and brightness have to be memorized for proper navigation.

The strength of mapping principle is that it is intuitive in interactive use. The further the user moves away from the target, the less pleasing [16, 26] the sound becomes: fast pitch changes sound siren-like, fast beats sound chopped, a high degree of roughness is an indicator of inharmonicity, low fullness sound like budget loudspeakers and a high degree of brightness sounds shrill. When moving away from the target, the user recognizes that the sound becomes more and more annoying and stressful.

3 METHOD

At the moment we are carrying out experiments with passive listeners. The experiment resembles the one that we have carried out to evaluate our two dimensional sonification [24, 25]. The results of the previous experiments demonstrated that these two dimensions are in fact readily interpretable and orthogonal. With the new experiment we aim at evaluating whether the new, third dimension (the z-dimension) is also orthogonal to the previous ones. If so, the direction and distance along the new dimensions should be interpretable even in the presence of the other two dimensions. To test this, we assign each participant to one of three groups. Each group is assigned to one of the three two dimensional pairs, i.e., x-y, x-z and z-y. This makes the results comparable to the results of the previous study. Furthermore, it is easier for users to learn and evaluate only two dimensions at a time. Experiments with three dimensions would take much longer, which is inefficient at this early stage of evaluation.

Each participant has the same task in two dimensional space,

but with different sonification metaphors. First, the sonification principles are explained to the participants, axis by axis. Then, the participants can explore each axis interactively, by moving a visible target along an axis with a computer mouse. After this explanation and exploration phase, which took about five minutes, the actual task is explained to the participants. We show them a map with 16 fields. The same map is also visible in the background of Figs. 2 to 4. The current location is the origin of the coordinate system. Each of the 16 fields lies either slightly to the left or right and up or down (the inner circle), far towards both orthogonal directions (the four outer fields), or slightly towards one direction and far towards the other (the ring in the middle). One sound is be played to them, representing an invisible target in either of the fields. The task of the participant is to interpret the sound and click on the target field with a cursor. No feedback is given. After clicking in one field, the next sound is played. We tell them that the order of fields was random, and that it was possible that some fields might not be sonified at all, sonified once, or even several times. However, we assured that each target would be played at least once to each participant. The participants are allowed to listen to the sonification as long as desired and adjust the volume according to their needs. Twenty sounds are played to each participant.

Group 1 is a reference group. They solve the task in the x-y-plane, illustrated in Fig. 2. The experiment resembles the experiment described in [24, 25]. Group 2 solves the task in x-z-plane, i.e., they tests the new dimension against the x-dimension as shown in Fig. 3. Group 3 solves the task in z-y-plane, i.e., they tests the new dimension against the y-dimension as as illustrated in Fig. 4.

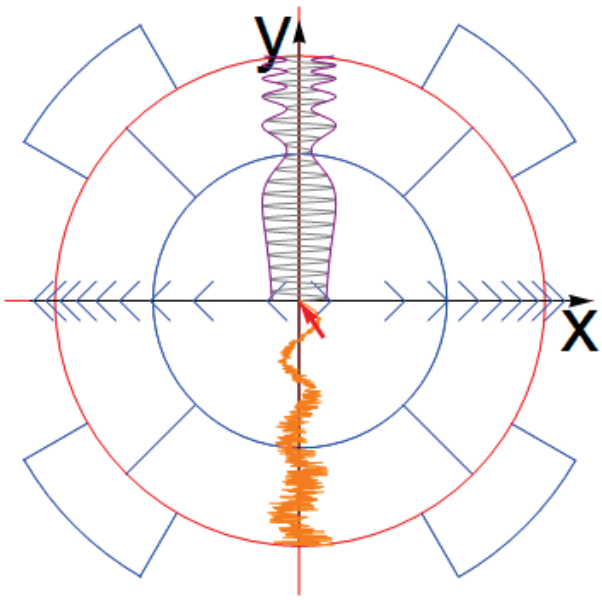


Figure 2: Sonification metaphors and map for group 1: the x-y plane.

So far we tested $N = 12$ participants. Each participant is assigned to one of three groups, i.e., we have 4 participants per group. The initial results of the ongoing study are presented and discussed in the following sections. Note, however, that the number of participants is too low to generalize the results and draw conclusions. In our previous experiment [24, 25] one out of 7 participants performed at chance level and we observed large inter-individual differences in other sonification studies as well [4, 27]. Hence, the reported initial results give a first indication of the suitability of the z-dimension, but they are not robust against outliers with exceptionally good or poor performance.

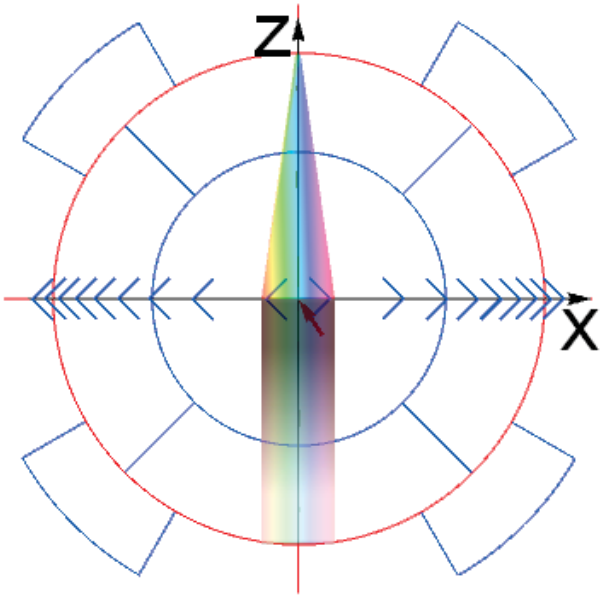


Figure 3: Sonification metaphors and map for group 2: the x-z plane.

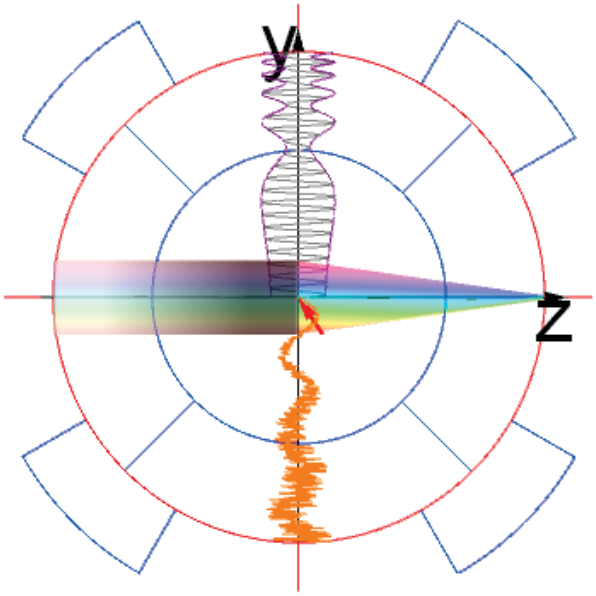


Figure 4: Sonification metaphors and map for group 3: the z-y plane.

Group	1	2	3
Plane	x-y	x-z	z-y
hits	72.5%	56.25%	53.75%
neighbor	20%	33.75%	36.25%
x-direction	95%	95%	—
y-direction	98.75%	—	97.5%
z-direction	—	93.75%	95%
quadrant	93.75%	88.75%	92.5%

Table 1: Results of the experiment per group.

4 RESULTS

The results of the experiment are summarized in Table 1. Obviously, participants in all groups performed much better than random guesses: All three groups hit the target field in over 53% of all trials. Chances to guess the correct field are $1/16 = 6.25\%$. In 20 to 36% of the trials the direct neighbor of the target field was selected by the participants. The correct direction along each single dimension was chosen in over 93% of all trials. The correct quadrant was chosen in 88 to almost 93% of all trials.

Overall, the performance of group 2 and 3 is in a similar range as group 1. One exception is the hit rate, which is much higher in group 1. However, the hit rate of group 1 is also much higher than in our previous study [24, 25], where the hit rate was 41%. The hit-rate of the individual users ranged from 30 to 90% in group 1, 15 to 85% in group 2, and 50 to 60% in group 3. This large variety of individual performances is evidence that experiments with a larger group are necessary to average out the exceptionally high or low performance of individual participants.

5 DISCUSSION

The results indicate that participants readily interpret each of the dimensions, even in the presence of a second dimension. In most cases they found the correct quadrant, i.e., they were able to distinguish the different sound characteristics well. This indicates that the sound characteristics are in fact approximately orthogonal in perception. The high hit-rate of over 53% indicates that the participants were also readily able to distinguish the two distances along each direction. This is a first indication of a high resolution. However, further experiments are necessary to evaluate the resolution of the dimensions.

If the results with more participants show a significantly higher hit rate in the x-y-plane than in the other planes, we will have to find out the reasons for that and act accordingly. If the results with more participants are promising, an interactive experiment as described in [8], and more thoroughly discussed in [4], will follow. If one or both directions of the z-dimension turn out to be uninterpretable during the presence of one or both directions of the x- or y-dimension, we will modify or replace the sonification of the directions accordingly. In [5] we already discussed options to modify the sonification. To our subjective impression, the fullness metaphor is the weakness of the three dimensional sonification. Depending on its implementation it is either not perfectly orthogonal to the beats or brightness directions, or it has a low resolution.

To this point it seems that we have successfully derived and implemented a monaural sonification for three dimensional navigation. After testing more participants we will be able to conclude if one or both of the new half-dimensions fulfill the requirements for three dimensional navigation or if further improvements are necessary.

References

[1] Thomas Hermann, Andy Hunter, and John G. Neuhoff, editors. *The Sonification Handbook*. COST and Logos, Berlin, 2011. URL http://sonification.de/handbook/.

[2] Gaëtan Parseihian, Charles Gondre, Mitsuko Aramaki, Solvi Ystad, and Richard Kronland-Martinet. Comparison and evaluation of sonification strategies for guidance tasks. *IEEE Trans. Multimedia*, 18(4):674–686, April 2016. ISSN 1520-9210. doi:10.1109/TMM.2016.2531978.

[3] Tim Ziemer, Holger Schultheis, David Black, and Ron Kikinis. Psychoacoustical interactive sonification for short range navigation. *Acta Acustica united with Acustica*, 104(6):1075–1093, 2018. ISSN 1610-1928. doi:10.3813/AAA.919273.

[4] Tim Ziemer and Holger Schultheis. Psychoacoustic auditory display for navigation: an auditory assistance system for spatial orientation tasks. *J. Multimodal User Interfaces*, Special Issue: Interactive Sonification, 2018. doi:10.1007/s12193-018-0282-2.

[5] Tim Ziemer and Holger Schultheis. Psychoacoustical signal processing for three-dimensional sonification. In *25th International Conference on Auditory Displays (ICAD2019)*, Newcastle, June 2019.

[6] Tapio Lokki and Matti Gröhn. Navigation with auditory cues in a virtual environment. *IEEE MultiMedia*, 12(2):80–86, April 2005. ISSN 1070-986X. doi:10.1109/MMUL.2005.33.

[7] Bruce N. Walker and Jeffrey Lindsay. Navigation performance with a virtual auditory display:effects of beacon sound, capture radius, and practice. *Human Factors*, 48(2):265–278, 2006.

[8] Tim Ziemer and Holger Schultheis. A psychoacoustic auditory display for navigation. In *24th International Conference on Auditory Displays (ICAD2018)*, Houghton, MI, June 2018. doi:10.21785/icad2018.007.

[9] David Black, Christian Hansen, Arya Nabavi, Ron Kikinis, and Horst Hahn. A survey of auditory display in image-guided interventions. *International Journal of Computer Assisted Radiology and Surgery*, 12(9):1665–1676, 2017. doi:10.1007/s11548-017-1547-z.

[10] Tim Ziemer and David Black. Psychoacoustically motivated sonification for surgeons. In *International Journal of Computer Assisted Radiology and Surgery*, volume 12, pages 265–266, Barcelona, Jun 2017. URL https://link.springer.com/article/10.1007/s11548-017-1588-3.

[11] Holger Schultheis, Tim Ziemer, and Tina Vajsbaher. Cognitive surgical assistance in minimally-invasive interventions. In *8. Interdisziplinären Workshops zu Kognitiven Systemen*, pages 12–13, Duisburg-Essen, March 2019. doi:10.17185/duepublico/4840.

[12] Douglas S. Brungart, Brian D. Simpson, Ronald C. Dallman, Griffin Romigh, Richard Yasky, and John Raquet. A comparison of head-tracked and vehicle-tracked virtual audio cues in an aircraft navigation taask. In *Proceedings of the 13th International Conference on Auditory Display*, pages 32–37, Montréal, Jun 2007. URL http://hdl.handle.net/1853/49972.

[13] André Lundkvist, Roger Johnsson, Arne Nykänen, and Jakob Stridfelt. 3d auditory displays for parking assistance systems. *SAE Int. J. Passeng. Cars – Electron. Electr. Syst.*, 10:17–23, 04 2017. doi:10.4271/2017-01-9627.

[14] He Huang, Steven L Wolf, and Jiping He. Recent developments in biofeedback for neuromotor rehabilitation. *Journal of NeuroEngineering and Rehabilitation*, 3(11), 2006. doi:10.1186/1743-0003-3-11.

[15] Daniel S. Scholz, Sönke Rohde, Nikou Nikmaram, Hans-Peter Brückner, Michael Großbach, Jens D. Rollnik, and Eckart O. Altenmüller. Sonification of arm movements in stroke rehabilitation – a novel approach in neurologic music therapy. *Frontiers in Neurology*, 7:Article ID: 106, 2016. ISSN 1664-2295. doi:10.3389/fneur.2016.00106.

[16] Eberhard Zwicker and Hugo Fastl. *Psychoacoustics. Facts and Models*. Springer, Berlin, Heidelberg, second updated edition, 1999. doi:10.1007/978-3-662-09562-1.

[17] D. Arfib, J.M. Couturier, Kessous, L., and V. Verfaille. Strategies of mapping between gesture data and synthesis model parameters using perceptual spaces. *Journal of Organised Sound*, 7(2):127–144, 2002. doi:10.1017/S1355771802002054.

[18] Jens Blauert. *Spatial Hearing. The Pychophysics of Human Sound Source Localization*. MIT Press, Cambridge, MA, revised edition, 1997.

[19] Albrecht Schneider. *Pitch and Pitch Perception*, pages 605–685. Springer Berlin Heidelberg, Berlin, Heidelberg, 2018. ISBN 978-3-662-55004-5. doi:10.1007/978-3-662-55004-5_31.

[20] Roger N. Shepard. Circularity in judgments of relative pitch. *The Journal of the Acoustical Society of America*, 36(12):2346–2353, 1964. doi:10.1121/1.1919362.

[21] Tim Ziemer. *Psychoacoustic Music Sound Field Synthesis (in Print)*, volume 6 of *Current Research in Systematic Musicology*. Springer, Cham, 2019.

[22] Albrecht Schneider. Perception of timbre and sound color. In Rolf Bader, editor, *Springer Handbook of Systematic Musiwology*, chapter 32, pages 687–726. Springer, Berlin, Heidelberg, 2018. doi:10.1007/978-3-662-55004-5_32.

[23] Tim Ziemer, Yi Yu, and Suhua Tang. Using psychoacoustic models for sound analysis in music. In *Proceedings of the 8th Annual Meeting of the Forum on Information Retrieval Evaluation*, FIRE '16, pages 1–7, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-4838-6. doi:10.1145/3015157.3015158.

[24] Tim Ziemer, David Black, and Holger Schultheis. Psychoacoustic sonification design for navigation in surgical interventions. *Proceedings of Meetings on Acoustics*, 30, 2017. doi:10.1121/2.0000557.

[25] Tim Ziemer. Two-dimensional psychoacoustic sonification. In Franziska Olbertz, editor, *33. Jahrestagung der deutschen Gesellschaft für Musikpsychologie (DGM)*, pages 60–61, Hamburg, Sep 2017. URL https://www.researchgate.net/publication/319778727_Two-dimensional_psychoacoustic_sonification.

[26] W. Aures. Berechnungsverfahren für den sensorischen wohlklang beliebiger schallsignale (a model for calculating the sensory euphony of various sounds). *Acustica*, 59(2):130–141, 1985.

[27] Tim Ziemer and Holger Schultheis. Perceptual auditory display for two-dimensional short-range navigation. In *Fortschritte der Akustik — DAGA 2018*, pages 1094–1096, Munich, March 2018. Deutsche Gesellschaft für Akustik. URL https://www.dega-akustik.de/publikationen/online-proceedings/.

Keynote Speech 2:
Prof. Dr. Rolf Bader
(Professor for Systematic Musicology at the Institute of Systematic Musicology, University of Hamburg)

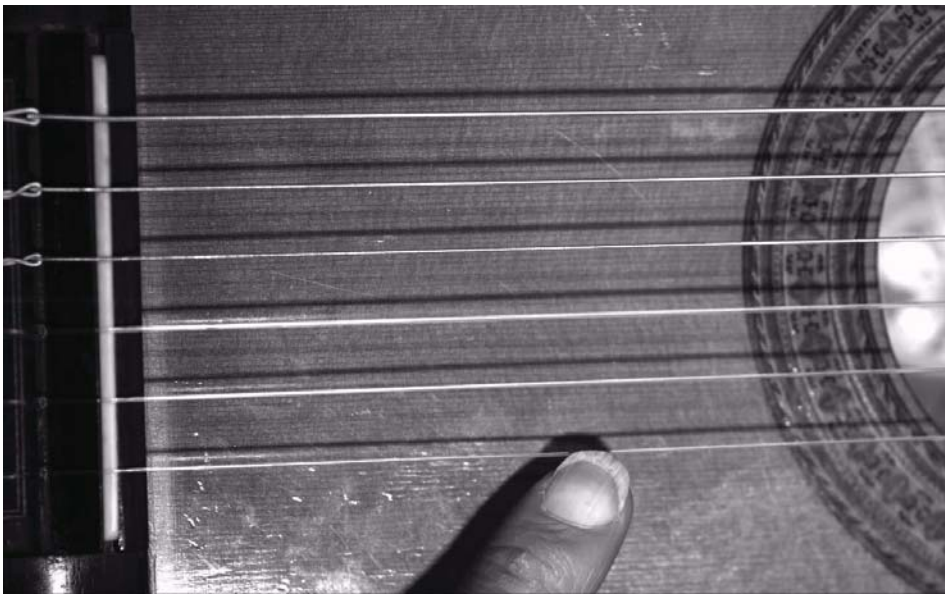
The future in physical modeling and auralization in music

Rolf Bader
Institute of Systematic Musicology
University of Hamburg

Overview

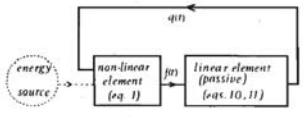
- Physical Models (Structural Mechanics, Fuid Dynamics, Room Acoustics, Viscoelasticity, Self-organization, etc.)
- Finite-Difference Time Domain (FDTD), Finite-Element Method (FEM), Impulse Pattern Formulation (IPF), etc.
- Hardware acceleration (FPGA, GPU)
- Artificial Intelligence (Neural Networks, Kohonen maps, etc.)
- Research question
- Manufacturer, Consumer demands

Gitarrensaite

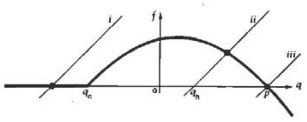


Impulse Pattern Formulation (IPF)

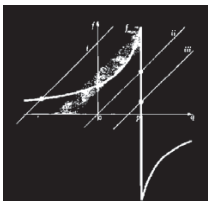
- Musikinstrumente werden fast alle durch Impulse beschrieben, die generiert und reflektiert werden
- McIntyre et al. 1983: Faltung eines Impulses mit nichtlinearer Generatorfunktion
- Fletcher 1985: Mode locking von Impulsen durch nichtlinearen Generator



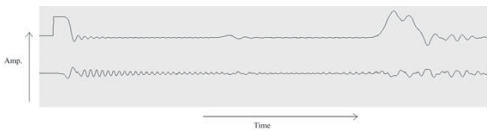
Feedback loop
Reed



Generator Saite/Bogen bei Geige



Generator



Impulse Train Guitar, one period

Impulse Pattern Formulation (IPF)

Two back impulses:

$$\frac{\partial g(t)}{dt} = -\alpha g(t) - \beta g(t)$$

Leads to:

$$-\frac{1}{\alpha} g(t) + \frac{\beta}{\alpha} e^{g(t)-g(t)} = e^{g(t)-g(t)}$$

Instantaneous:

$$g(t) = g(t) - \ln\left[\frac{1}{\alpha} g(t) - \frac{\beta}{\alpha} e^{g(t)-g(t)}\right]$$

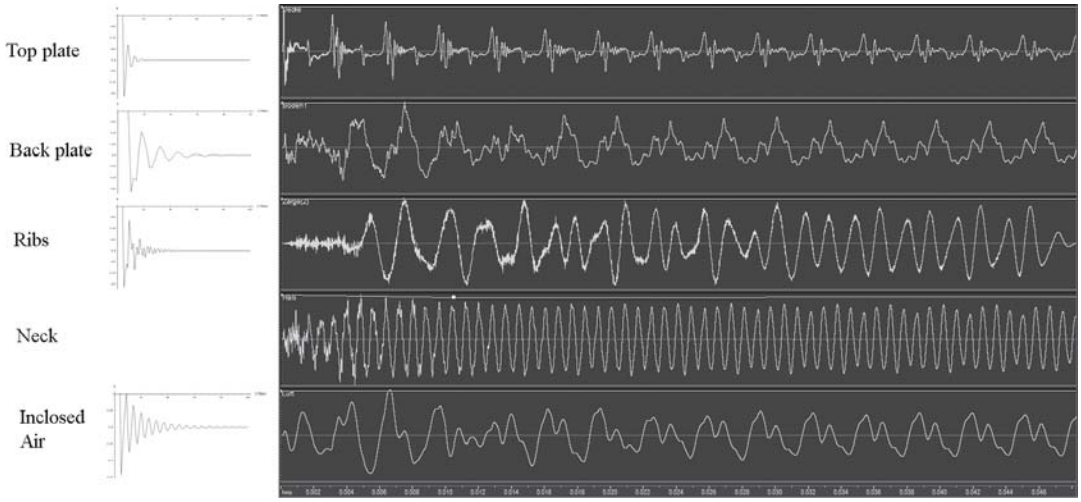
Delayed:

$$g(t_+) = g(t) - \ln\left[\frac{1}{\alpha} g(t) - \frac{\beta}{\alpha} e^{g(t)-g(t_-)}\right]$$

n back impulses
Delayed:

$$g(t_+) = g(t) - \ln\left[\frac{1}{\alpha} g(t) - \sum_{k=1}^{n} \frac{\beta_k}{\alpha} e^{g(t)-g(t_{-k})}\right]$$

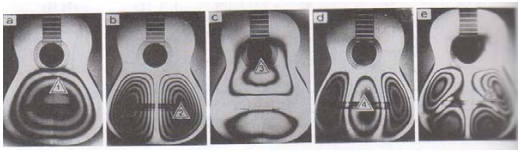
Guitar Parts



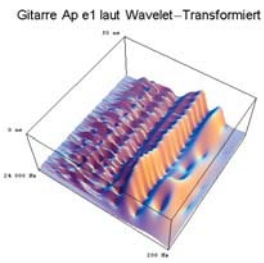
Microphone Array

128 microphones in an array of 11 x 11 lines for nearfield recording of structures in a distance of 3 cm – 10 cm. Estimation possible for surface description and radiation character differences of different surface points, location of far-field sources, diffusion estimation and of course radiation measurements.

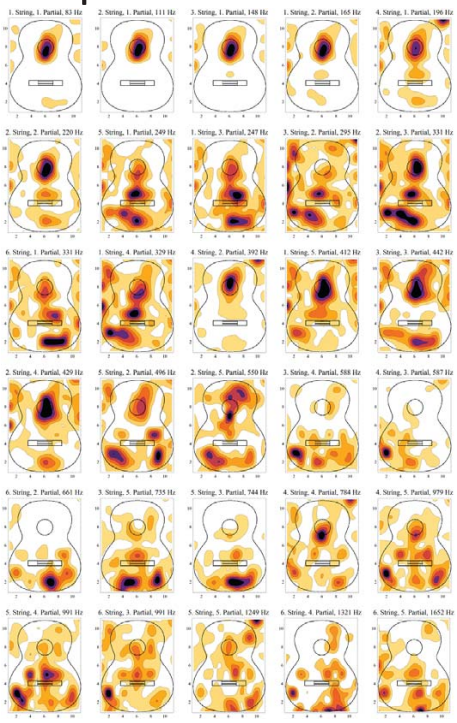
Strong radiation at spectral non-resonanc



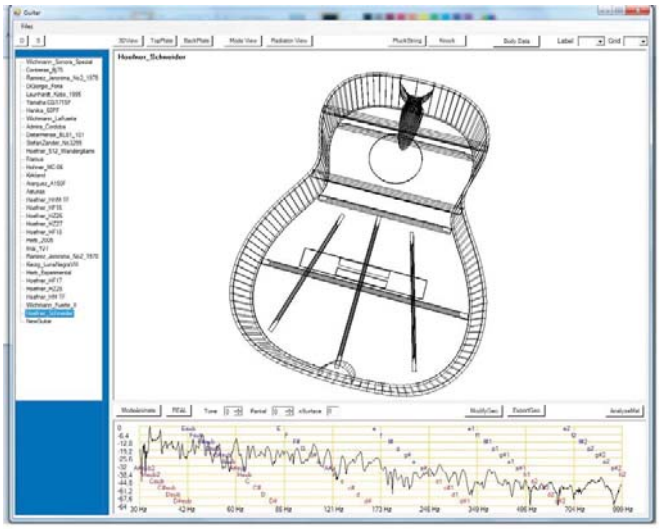
Jansson 1971



Spectral flux



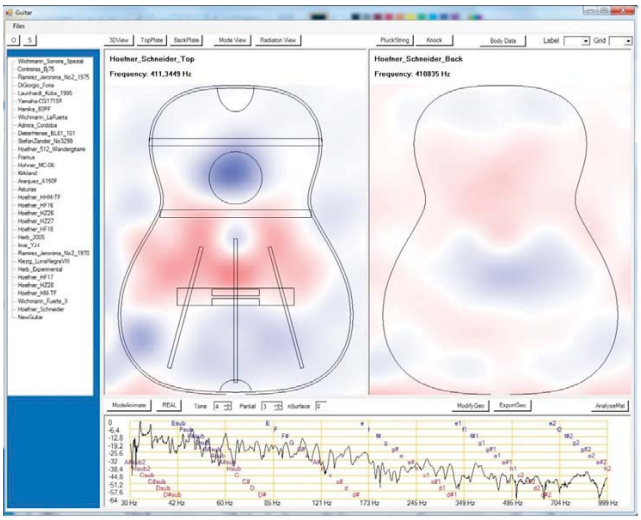
Digital Guitar Workshop Software



Screenshot of 3D view of our guitar building software

A computer program has been written to provide our conclusions to professional instrument makers. It will be able to compute sounds out of geometry changes and geometries out of sound spectra changes. This represents an inverse problem which is to be solved.

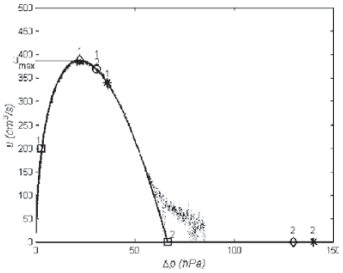
Digital Guitar Workshop Software

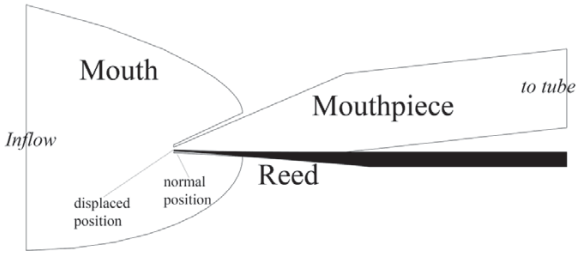


Screenshot of Mode View function of our guitar building software

Our projects target is to frame an answer on all constructional questions which influence the sound of classical, nylon stringed guitars, to show the whole extent of interaction between the measured parameters and to determine which one to change to achieve any aural imagination of guitar makers and players.

Saxophone mouthpiece

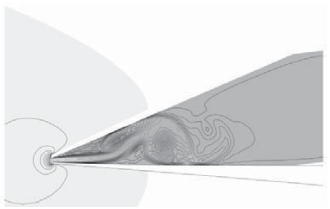




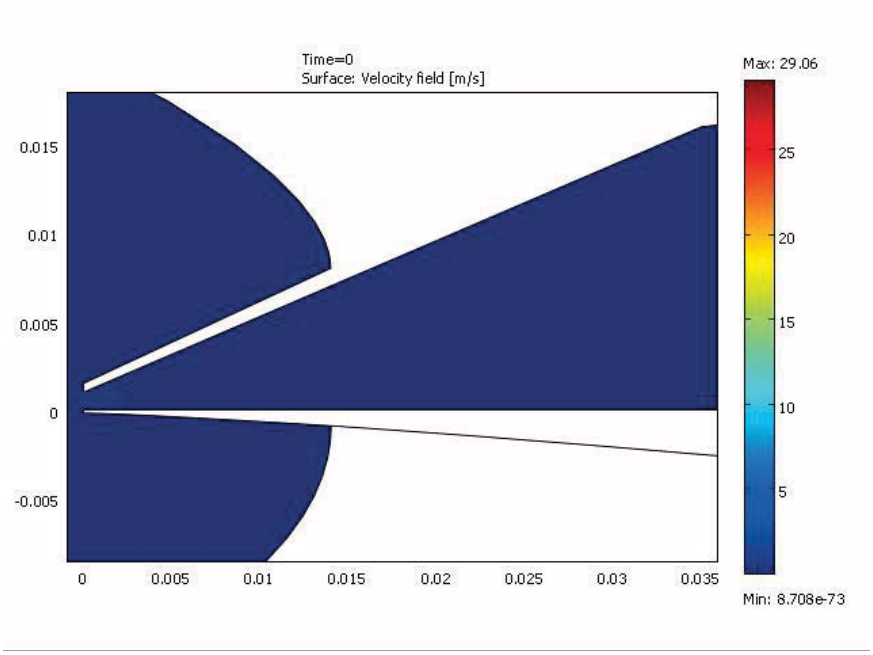
Druck vs. Fluß Saxophon, Messung und Rechnung

$$u = u_A \left(1 - \frac{\Delta p}{p_M}\right) \sqrt{\frac{\Delta p}{p_M}}$$

$$u_A = w H_0 \sqrt{\frac{2 K H_0}{\rho}} = \frac{3\sqrt{3}}{2} u_{max}$$

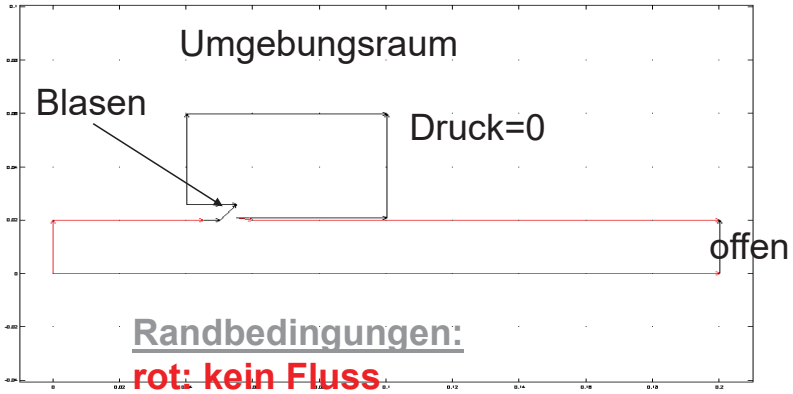


u: Fluß, p_M: Druck im Mund, p: Druckdifferenz Mund-Mundstück, w: Weite Mundstück, ρ: Dichte Luft, H_0: Öffnung Mundstück im Ruhezustand, K: Steifigkeit Rohrblatt



Turbulece in flutes

3.5 %



Randbedingungen:
rot: kein Fluss
schwarz: kein Druck

Reynolds-Averaged Navier-Stokes Equation

$$\partial_t \hat{u}_i + \hat{u}_j \partial_j \hat{u}_i = -\frac{1}{\rho} \partial_i p + \nu \nabla^2 \hat{u}_i \qquad i = 1, 2, 3 \qquad \hat{u}_i : \text{Flow} \qquad p : \text{Pressure}$$

$$\partial_t \hat{u}_i = 0 \qquad \rho : \text{Density} \qquad \nu : \text{Viscosity}$$

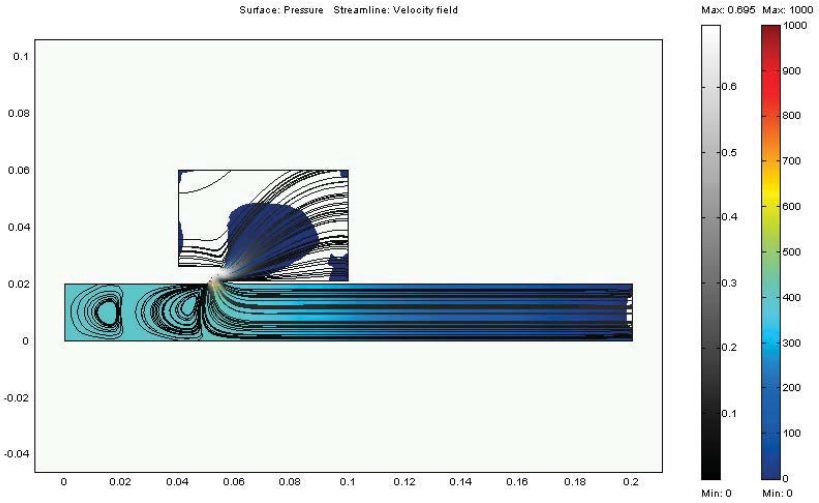
$$\hat{u} = U + u \qquad U : \text{mean velocity} \qquad u : \text{flucuta}$$

$$\partial_t U_i + U_j \partial_j U_i = -\frac{1}{\rho} \partial_i P + \nu \nabla^2 U_i - \partial_j \overline{u_j u_i}$$

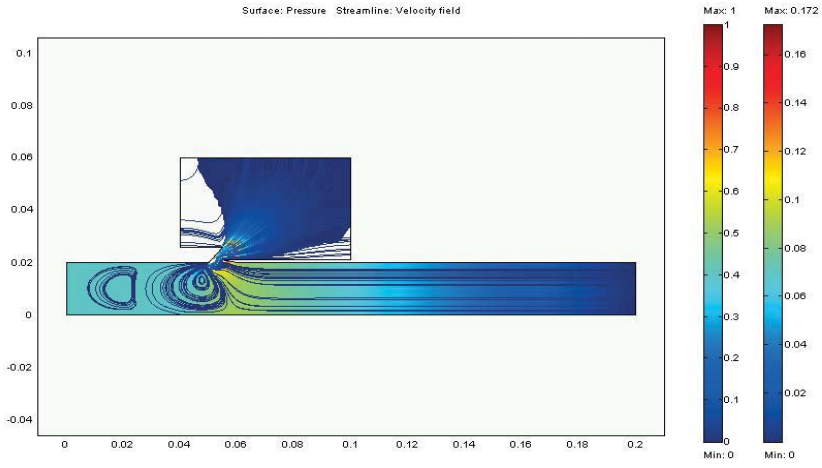
$$\partial_t U_i = 0$$

$$\partial_t \overline{u_i u_j} + U_k \partial_k \overline{u_i u_j} = -\frac{1}{\rho} \left(\overline{u_j \partial_i p} + \overline{u_i \partial_j p} \right) - 2\nu \overline{\partial_k u_i \partial_k u_j} - \partial_k \overline{u_k u_i u_j} - \overline{u_j u_k} \partial_k U_i - \overline{u_i u_k} \partial_k U_j + \nu \nabla^2 \overline{u_i u_j}$$

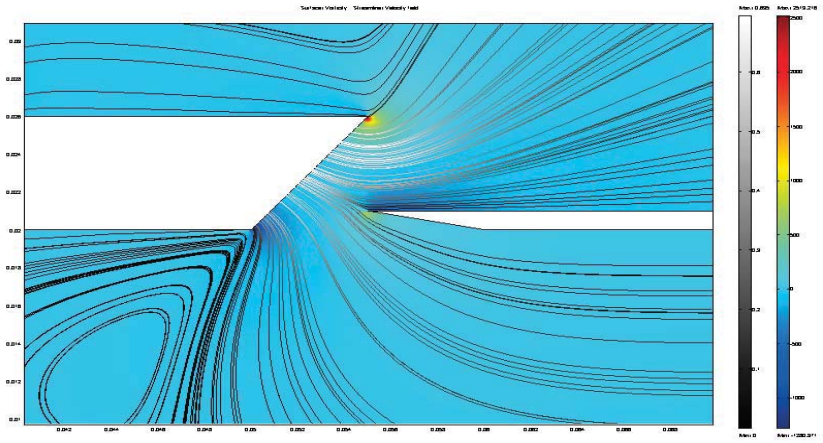
Navier-Stokes Modell



Turbulenz Modell

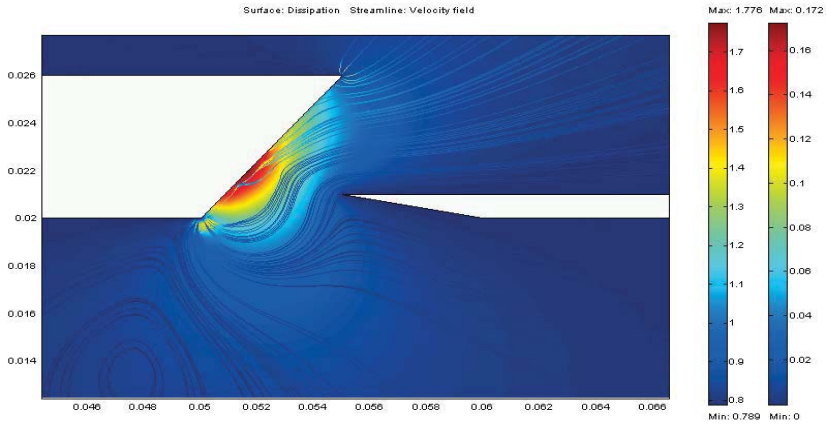


Navier-Stokes-Modell (Blasloch)



Turbulenz Modell (Blasloch)

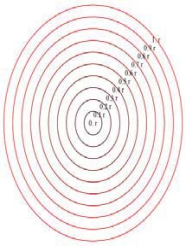
1.27 %



Tuning by adding ‚**pa sa**‘: paiste of (sticky) rice and ashes

Modes when changing mass in radius

- Doubling of mass over region
- 10 cases of additional mass: r = 0, r = 0.1, r = 0.2 .... r = 1
- 10 Modes considered

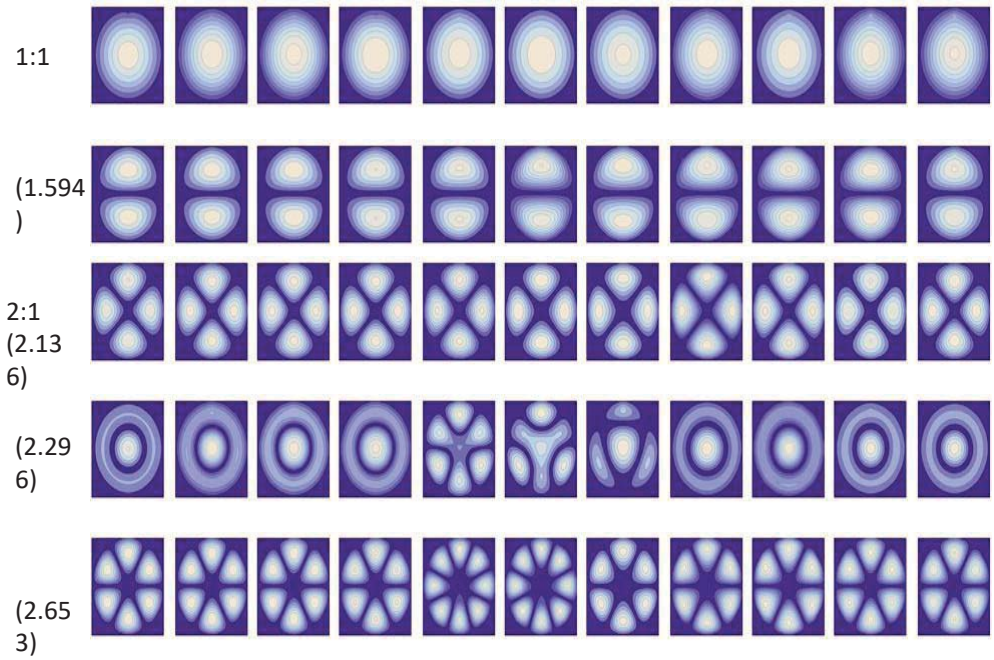


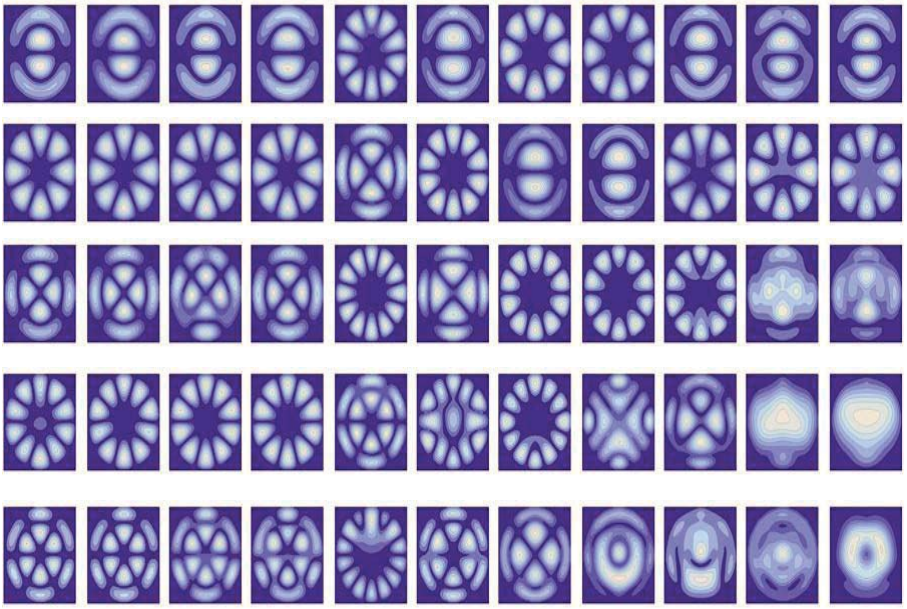
Membrane equation:
Radius r
Tension T
Area density μ
Mass m, Area A
Displacement u

$$\frac{T}{\mu(x,y)}\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}\right)=$$

$$\mu(x,y)=m(x,y)/A=\begin{cases}\mu_0 \text{ for } \sqrt{(x-x_0)^2+(y-y_0)^2}>0.1\,n\,r\\ 2\,\mu_0 \text{ for } \sqrt{(x-x_0)^2+(y-y_0)^2}<0.1\,n\,r\end{cases}$$

with n = 0,1,2,...10.





Sound examples of model Front and back membrane, Air modeled as delay

No viscosity-damping, front membrane

No viscosity-damping, back membrane

With viscosity-damping, front membrane

With viscosity-damping, back membrane

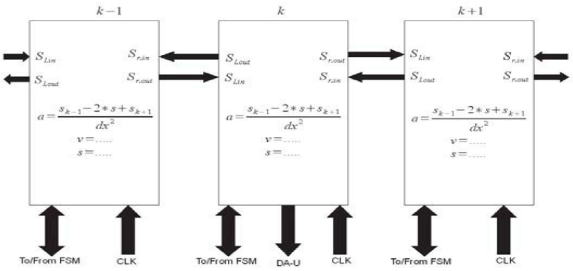
With viscosity-damping, pitch front membrane = pitch back membrane

With viscosity-damping, pitch front membrane < pitch back membrane

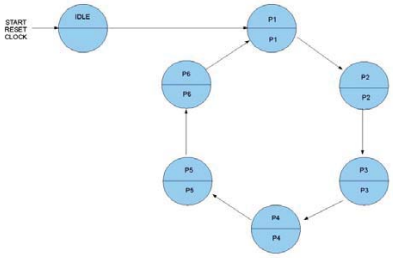
With viscosity-damping, pitch front membrane > pitch back membrane

Field-Programmable Gate Array (FPGA)

- Mixture of CPU and Memory Chip
- Massive Parallel Computation
- Serial Computation implementable (Finite-State Machine)
- Explicit Finite-Difference Model
- *Variable Data Type (VDT) arithmetics for real-time computation of whole musical instrument geometries*
- Input/Output on PC via PCI Express bus



Finite-Difference Hardware implementation.

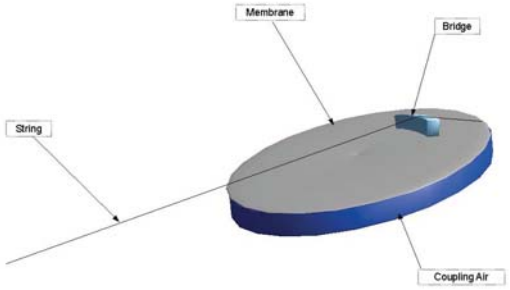


Finite-State Machine for serial computation.

Violin Model Sounds

- Violin top plate thickness change
- Different violin attacks
- Different playing positions
- Plucked
- Hammered

Banjo real-time FPGA Model

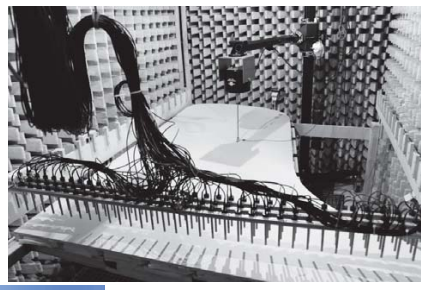


Banjo model with strings, bridge, membrane, and inclosed air.

Benjo Model Sounds

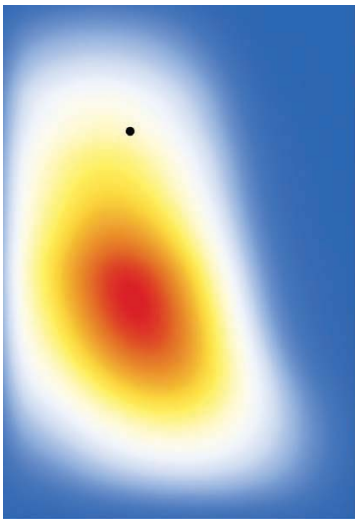
- Banjo with Finger-Pick model
- Banjo with varying membrane tension
- Banjo with varying low pitch

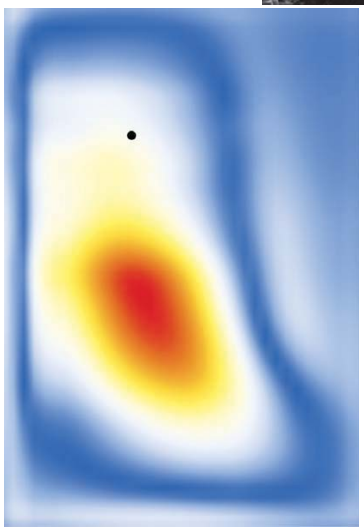
Precision of back-propagation



Raw data

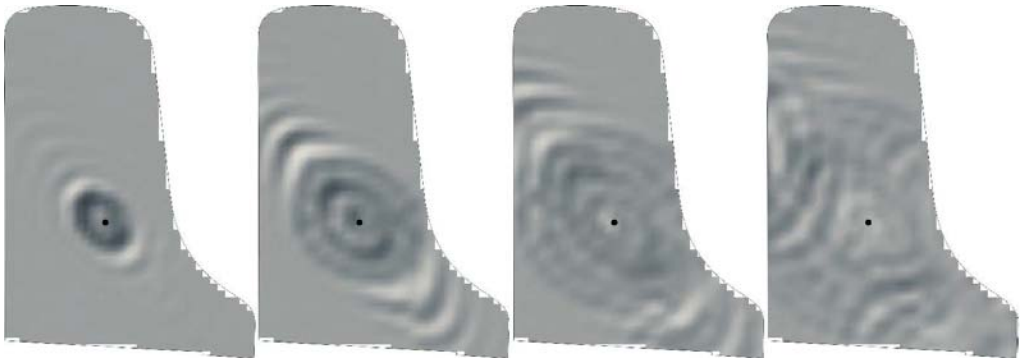
propagated to soundboard surface



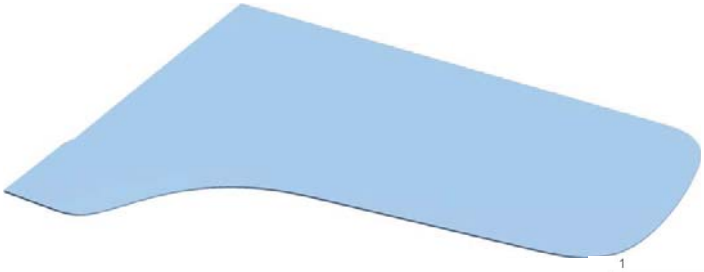


Piano concert grad recorded with 1800 microphones

Wave propagation of piano soundboard, 1-5ms



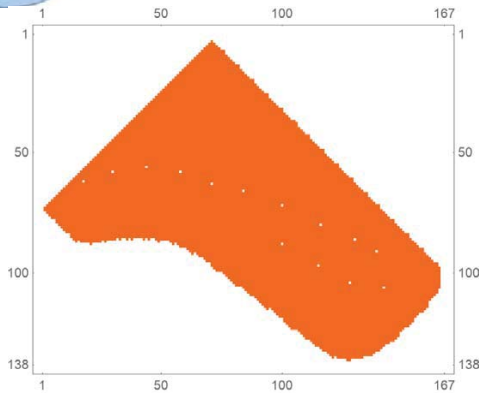
Physical Model



- Plate equation
- Fixed boundary conditions
- 14 driving points

- Sinusodal driving
- Impulse driving

- Implemented on 2 GPUs using CUDA programming language



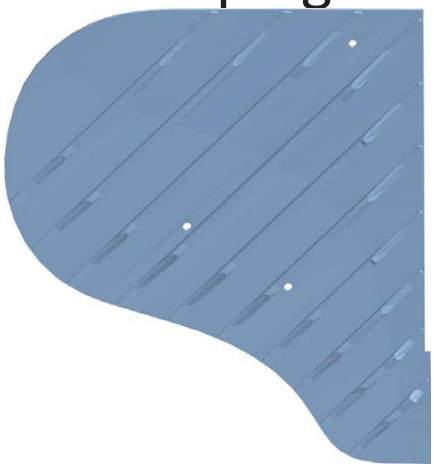
Importance of internal damping

Physical Model of Steinway M piano
Single string, with normal damping

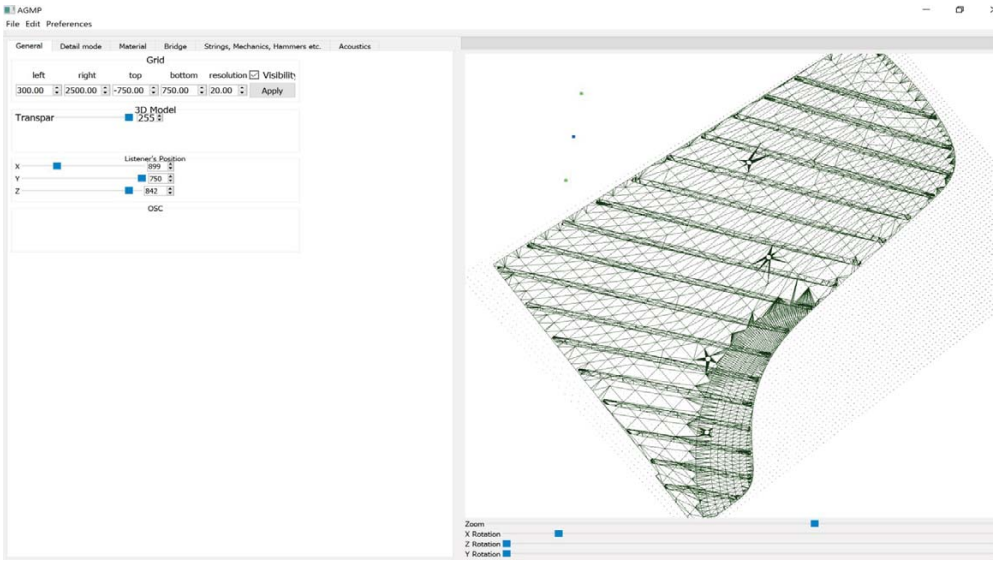
1) Normal soundboard damping with one parameter for all frequencies

2) Extreme soundboard damping

3) No soundboard damping



GUI for FPGA physical modeling piano design



Examples

Drum with long fundamental:

with viscosity

no viscosity

Piano soundboard (Steinway M-Model) knocking:

no viscosity

with viscosity 125 Hz damped 250 Hz damped 1000 Hz damped

Piano soundboard one string played:

no viscosity

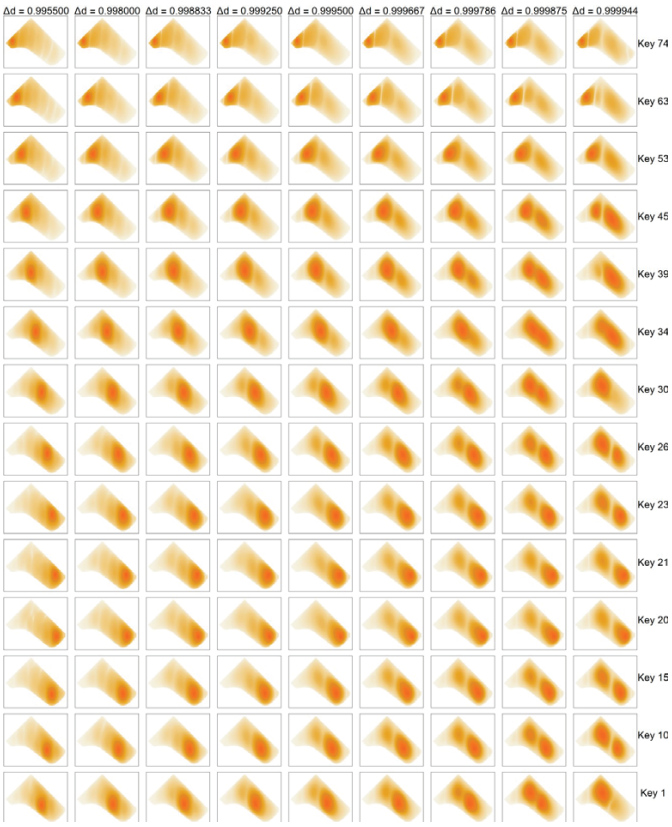
with viscosity

Eigenfrequency: **26 Hz**

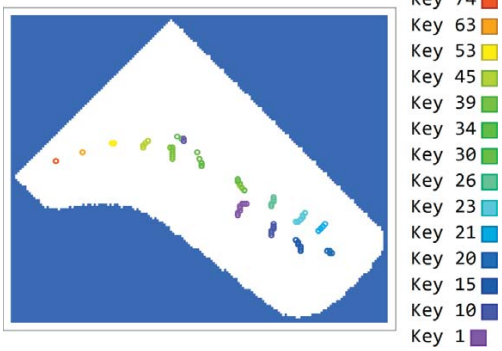
Sinusodal driven at 26 Hz

9 damping strength

14 driving points



Points of maximum energy



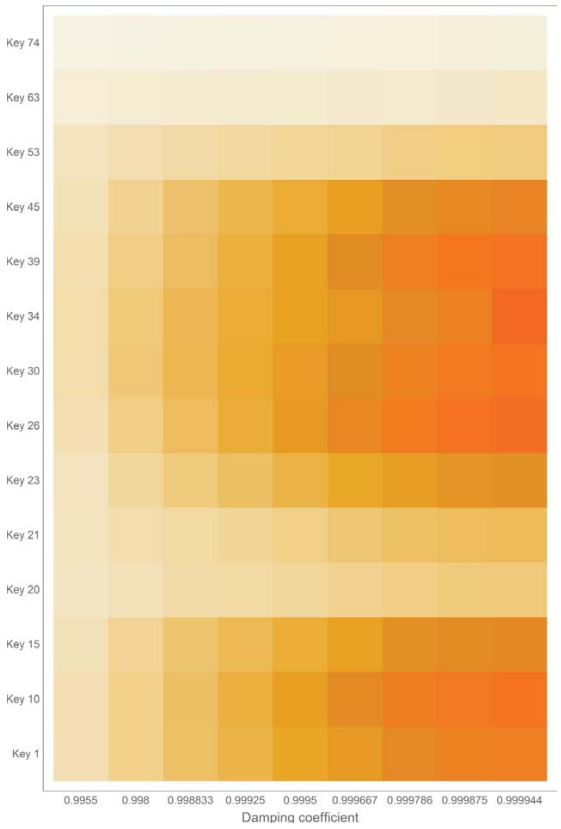
Eigenfrequency: **26 Hz**

Sinusodal driven at 26 Hz

9 damping strength

14 driving points

Energy integrated over plate



Leather glass transition damping

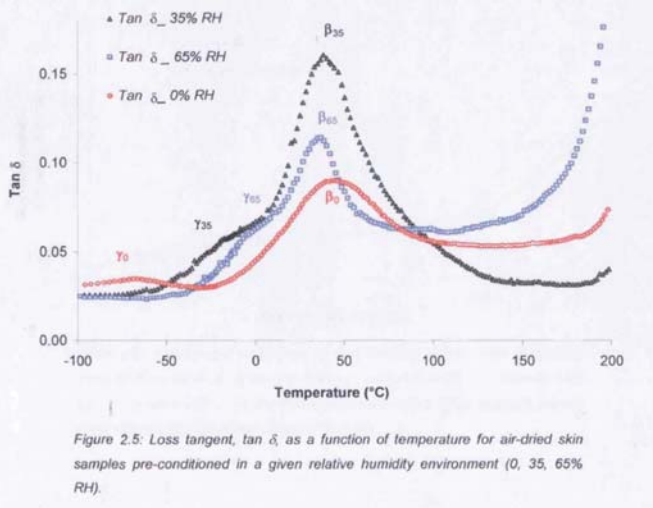


Figure 2.5: Loss tangent, tan δ, as a function of temperature for air-dried skin samples pre-conditioned in a given relative humidity environment (0, 35, 65% RH).

From: Sujeevini Jeyapalina. Studies on the hydro-thermal and viscoelastic properties of leather. PhD, Leichester 2004.

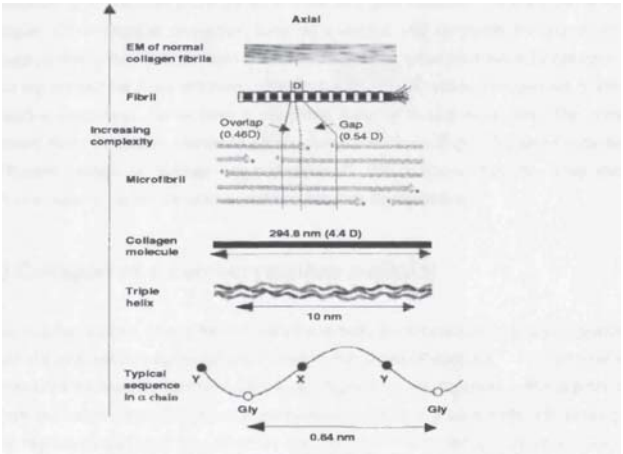


Figure 1.12: Schematic hierarchy of collagen structure, from alpha chain through to collagen Fibril[33].

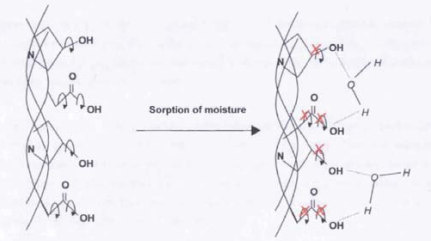
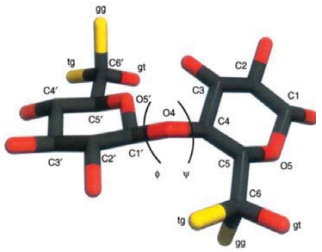


Figure 2.7: Schematic illustration of how the free rotation of hydrophilic side chain amino acids could be prevented when moisture is absorbed into the inter-fibril region.

Models of internal damping

- Thermal losses (Zener, Harris)
- Molecular conformational changes (Glassstone, Laidler, Eyring)



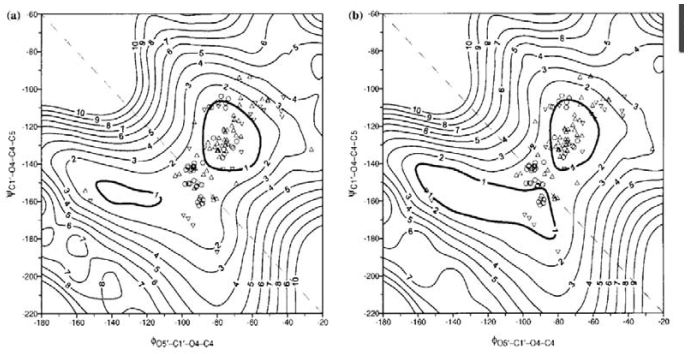


Figure 6. Energy surfaces for cellobiose at the (a) HF/6-31G(d) and (b) HF/6-311 + G(d) levels, based on the optimized HF/6-31G(d) geometries. Details are as in Figure 3.

From: Alfred D. French and Glenn P. Johnson: Advanced conformational energy surfaces for cellobiose. Cellulose 11: 449–462, 2004.

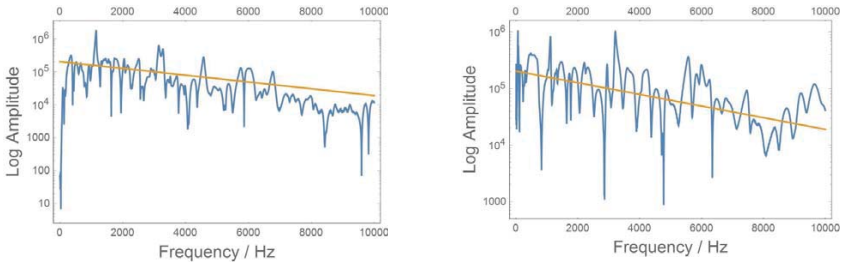
Kinds of damping

One damping parameter

$$\frac{T(x,y)}{\mu}\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}\right)=\frac{\partial^2 u}{\partial t^2}+D\,\frac{\partial u}{\partial t}$$

Frequency-dependent damping spectrum

$$\int_{\tau=0}^{T} h(\tau)\frac{T(x,y)}{\mu}\left(\frac{\partial^2 u(x,y,t-\tau)}{\partial x^2}+\frac{\partial^2 u(x,y,t-\tau)}{\partial y^2}\right)=\frac{\partial^2 u(x,y,t)}{\partial t^2}.$$



Spectra when knocking on wooden plate 30x30cm (left) before and (right) after laquering

Complex Young‘s modulus

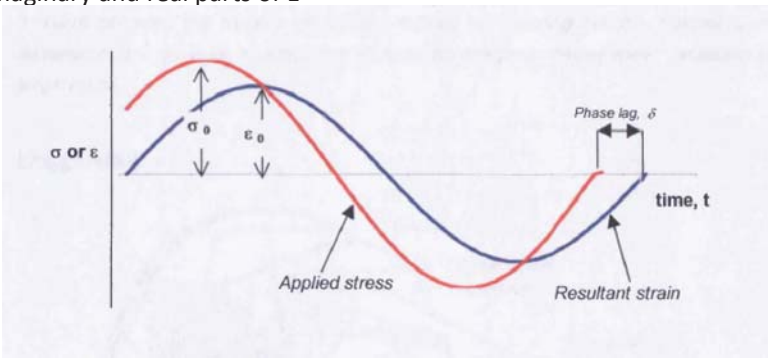
Complex stress-strain model:
σ : stress
ε: strain

$$\sigma(s) = E(s)\ \epsilon(s)$$

Phase-shift between stress and strain:

$$\tan\delta = E_I / E_R$$

E_I, E_R: imaginary and real parts of E



Model

Complex stress-strain model:
σ : stress
ε: strain
α: damping
ω: frequency
s: complex frequency

$$\sigma(s) = E(s)\ \epsilon(s)$$

$$s = \alpha + \imath\omega\ .$$

u: displacement
A: Amplitude
μ: damping constant

$$u(s,t) = A(s)\ e^{-\mu\ t}\ e^{\imath\ \omega\ t}$$

Transfer (1) into time domain (multiplication in frequency domain -> convolution in time domain)

$$\sigma(t) = \int_0^\infty \epsilon(t-\tau)\ h(\tau)\ d\tau\ .$$

Inverse Laplace transform:
γ: integration constant needed for convergence, influences damping strength Ds

$$h(\tau) = \frac{1}{2\pi\imath} \int_{s=\gamma-\imath\infty}^{s=\gamma+\imath\infty} E(s) e^{s\tau} ds$$

Model

Membrane
Differential Equation
u: displacement
T(x,y): tension
μ: area density
D: damping coefficient

$$\frac{T(x,y)}{\mu}\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}\right)=\frac{\partial^2 u}{\partial t^2}+D\,\frac{\partial u}{\partial t}$$

Viscoelastic membrane differential equation:

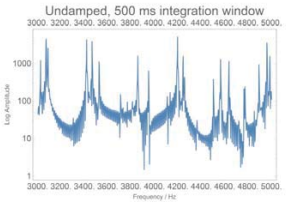
$$\int_{\tau=0}^{\infty} h(\tau)\frac{T(x,y)}{\mu}\left(\frac{\partial^2 u(x,y,t-\tau)}{\partial x^2}+\frac{\partial^2 u(x,y,t-\tau)}{\partial y^2}\right)\,d\tau\;=\frac{\partial^2 u(x,y,t)}{\partial t^2}$$

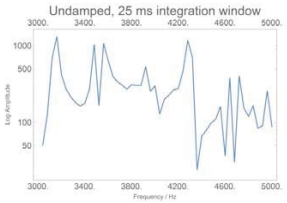
Computation time

GPU	N	Sound length	Computation time
GTX 940	1	1 sec	~ 20 sec
GTX 940	1000	1 sec	~ 380 sec
GTX 1070	1	1 sec	~ 12 sec
GTX 1070	1000	1 sec	~ 60 sec

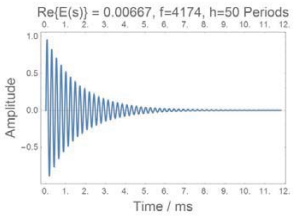
Result Target f = 4174 Hz

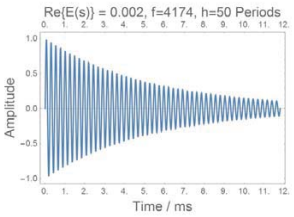
Length of h:
- 10 periods of f
- 25 periods of f
- 35 periods of f
- 50 periods of f

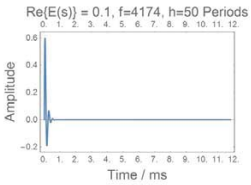




Each case γ = n x [10: 1/(n 10) ; 25: 1/(25 n) ; 35: 1/(25 n); 50: 1/(50 n)] with n = 1, 2, 3, … 10
Each case Re{Ek} = n x 0.0003 with n = 0, 2, 3, … 10

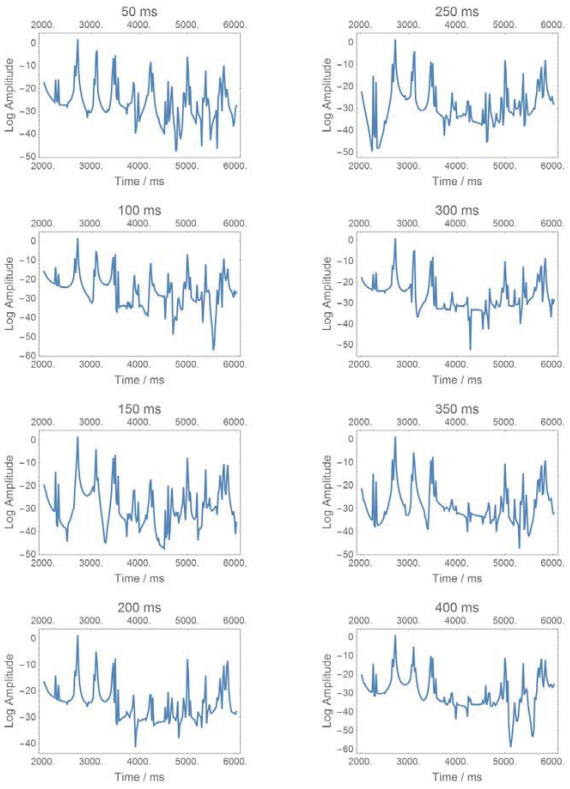




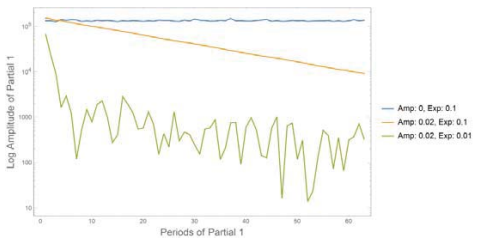


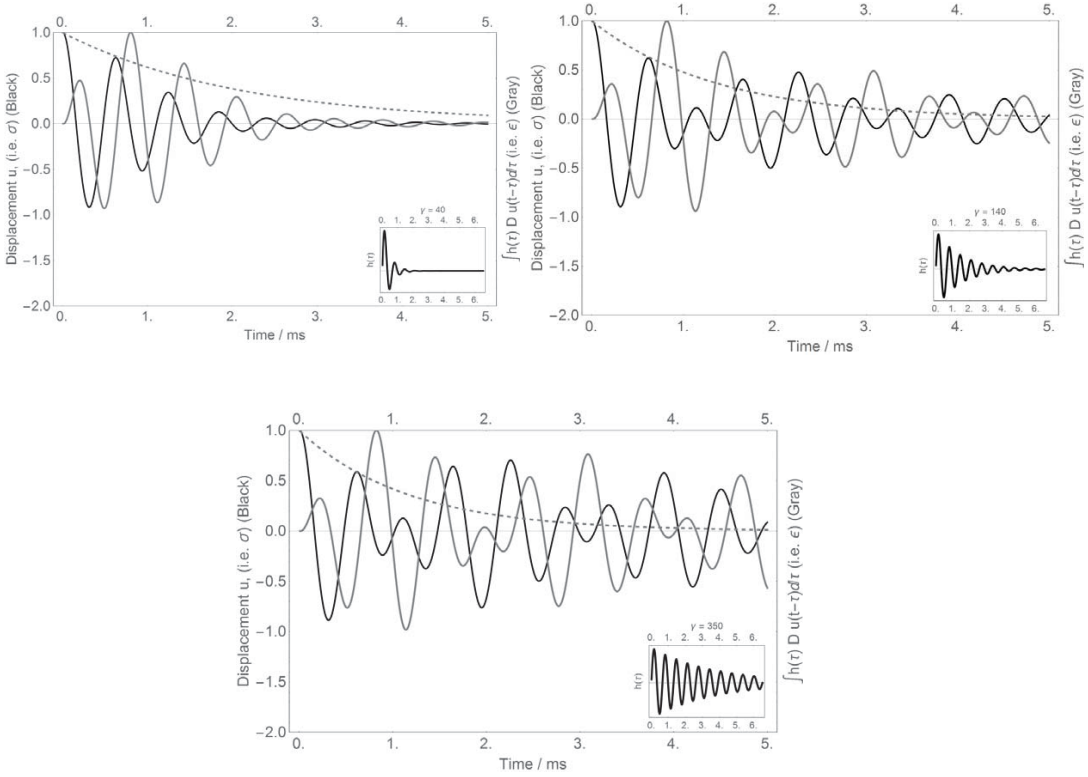
Damping of f = 4174 Hz

Example:
50 periods
Re{E(s)} = 1 / 0.0003

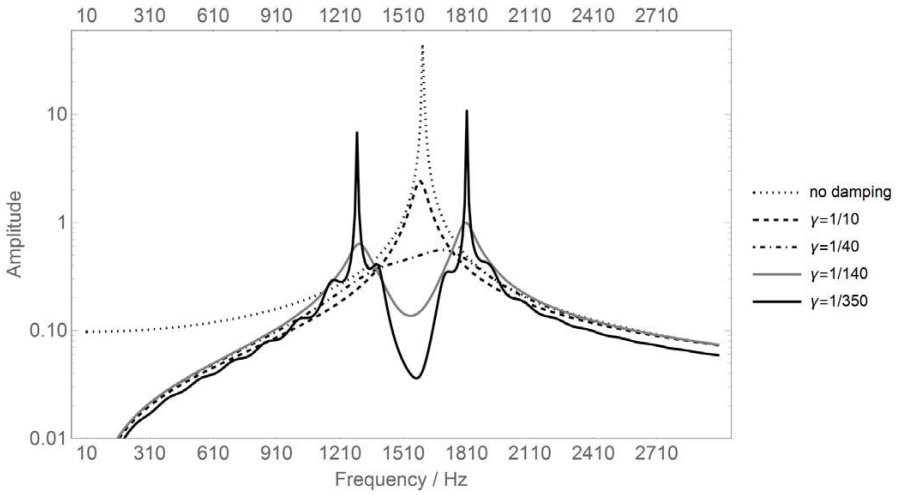


Partial decay μ from simulation





Mass point, Spetra



Non-exponential decay leads to sidebands with high γ.

Computational Phonogram Archiving

Music Information Retrieval
Networked Music Performance
Feature Extraction

- Onset detection (Lartiott & Toiviainen 2007 (MoIR Toolbox), Klapuri 2004, Dixon 2006 (review), Alexandraki & Bader 2014, …)
- Rhythm and meter (Dessein, Cont, and Lemaitre 2010, Goto2001, Klapuri 1999, Blass 2015, …)
- Pitch (Six 2006 (TARSOS), Goto 2001b, Klapuri & Davy 2006, Kob 2004, Chai & Vercoe 2003, Bader 2011, …)
- Score extraction (Benetos & Dixon 2013, Marchand 2015, …)
- Tonality (Leman & Carreras 1997, Kostek 2005, Bader 2013 (review), …)
- Timbre (Toiviainen , Bader 2013, …)
- Density (Gibiat & Casellengo 2000, Bader 2013, …)
- Musical Form (Foote 2000, Logan 2000, Foot 2003, Peiszer 2007, …)
- Speech estimation (Rabiner & Huang 1993, …)

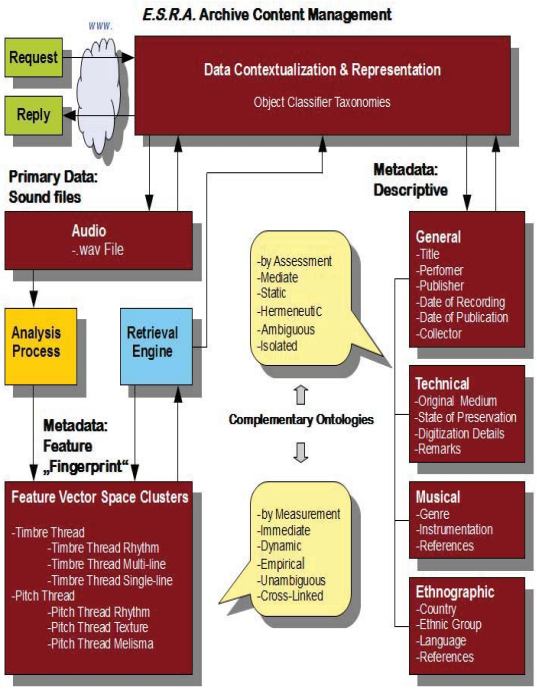
Methods

- MIR tools (Fourier, Cepstrum, Wavelet, Fractal Dimensions, Spectral Centroid, Roughness, Autocorrelation, Autoregression, …)
- Hidden Markov Models and Mixed Gaussian Models (Benetos & Dixon 2013 , Aucouturier & Sandler 2001, Peters et al. 2002, Levy et al. 2006, Alexandraki & Bader 2013, …)
- Self-organizing maps, Kohonen maps (Leman & Carreras 1997, Kostek 2005, Blass 2015, Toiviainen 1996…)
- Musical Acoustics and Physical Modeling (Papadiogiannis (lyra), Pfeifle & Bader 2012 (ruan), Hwang & Suzuki 2016 (Japanese drum), …)
- Microphone Array and Laser interferrometry (Nickerson & Rossing (Karen bronze drum), Bader 2012 (Iounuet), Rossing 2008, …)
- Wave Field Synthesis (Ziemer & Bader 2017, …)
- Internet Platform for Computational Ethnomusicological Archives (Franke 2016, …)
- Artificial Reality (Plath, Kirsch)

Application (selection)

- Computational, Analytical and Comparative Ethnomusicology (Louhivuori, Tilley, Azadeh)
- Archives (Place, Smithsonian Archive, Dorsch, African Music Archive, Koehn, ESRA Hamburg)
- Tonal systems of African xylophnes (Six)
- Organology (Sturm)
- Score extraction (Benetos, Maynard)
- Networked Music Performance (Alexandraki)
- Tuning of Gamelan ensembles (Wendt)
- Music Information Retrieval (Lukashevich, Fraunhofer Institut Illmenau, Da Cruz Lima, Simmone, Gestion et valorisation des archives du CREM)
- Musical Acoustics (Papadiogiannis, Kob, Pfeifle, Bader) Tonal system and scale detection (Flamenco, Raga, Carnatic music) (Compumusic, Serra)

Fig. 2: Schematic Representation of Database Core Ontologies



Feature Structure

Timbre

Kohonen map of multi-line timbre (TTP)

Kohonen map of single-line timbre, musical instrument sounds (TTH)

Pitch

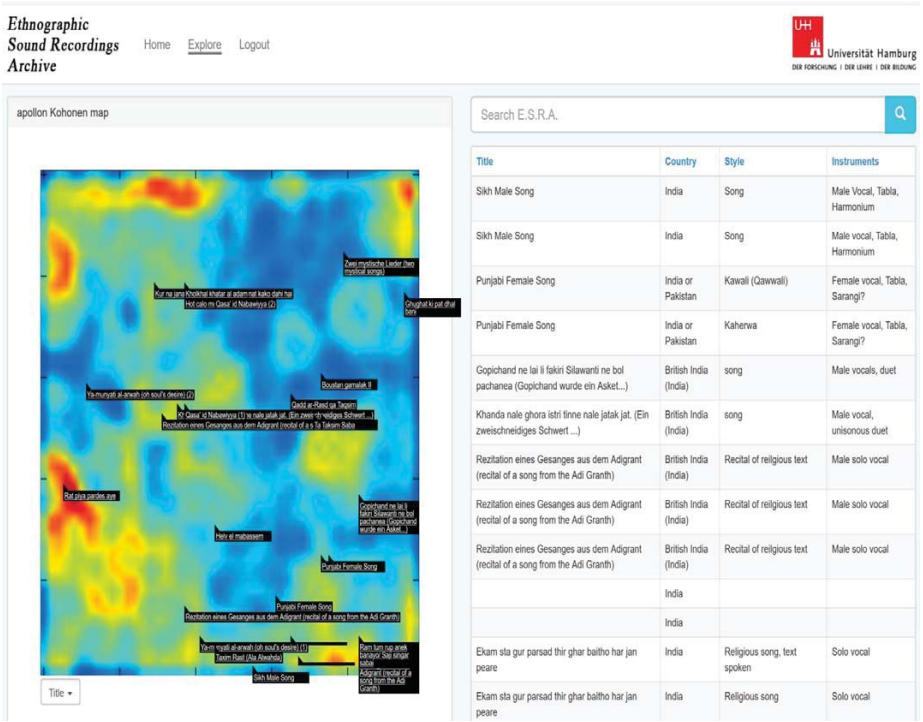
Score for multi- and single-line objects (PTT)

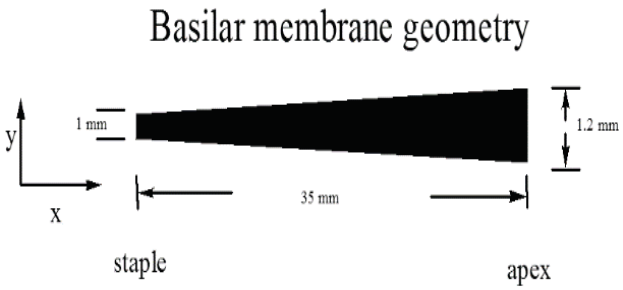
Kohonen map of melisma for all segments of all objects (PTM)

Rhythm

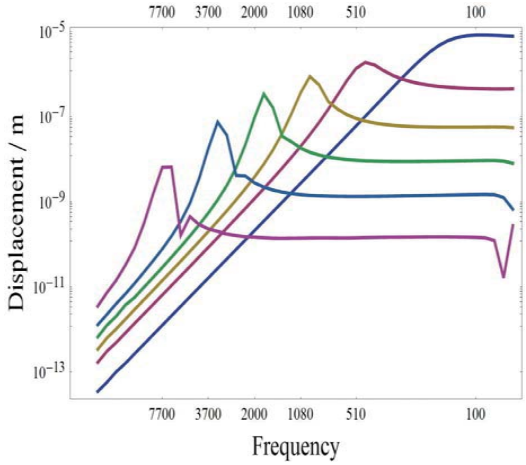
Hidden-Markov Model of multi-line fused sounds based on sound level (TTR)

Hidden-Markov Model of multi- and single-line objects based on pitch level (PTR)





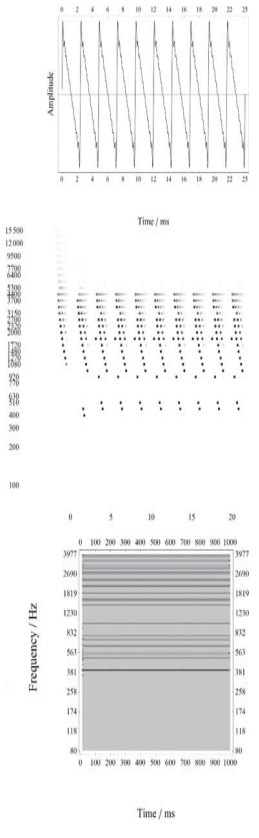
Best frequencies obtained from single sinusodial driving



10 partials

400 Hz f0

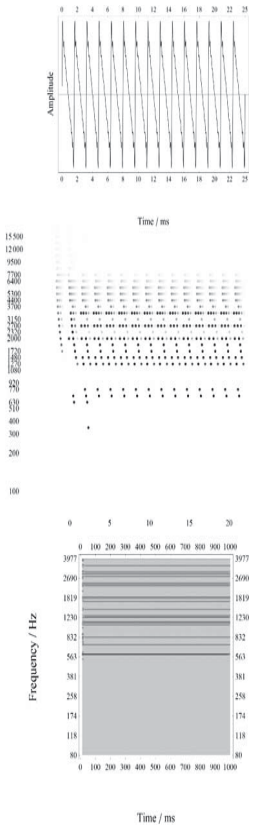
=> Lowest Periodicity: 400 Hz



10 partials

600 Hz f0

=> Lowest Periodicity: 600 Hz

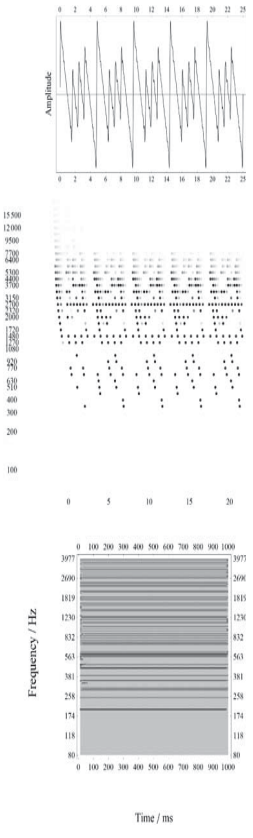


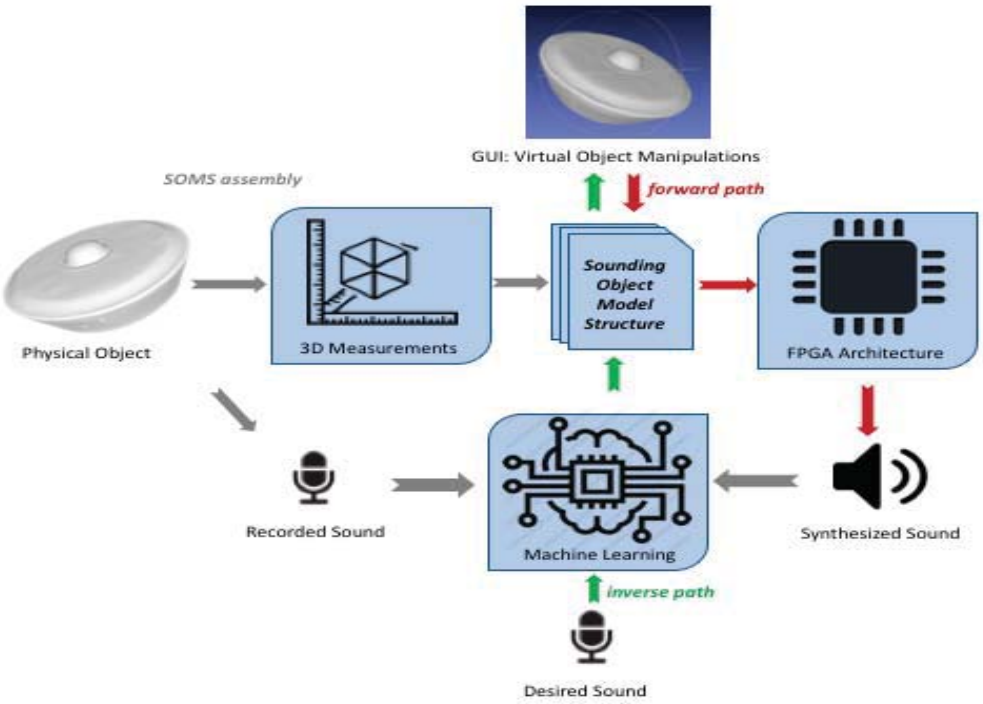
Two sounds:

10 partials each

400 Hz f0 + 600 Hz f0

=> Lowest Periodicity: 200 Hz





Conclusions

- Damping much more important for musical instrument sounds than modes
- Metamaterials offer promissing new methods for acoustical instruments
- Only FPGAs allow real-time implementation of real-geometry physical models
- Applications: instrument builders, musicians, phonogram archives, etc.

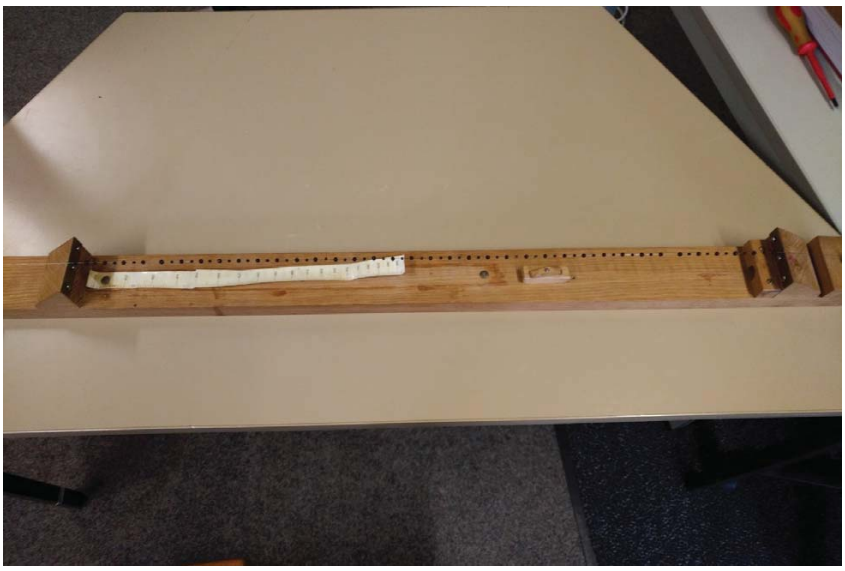
Musical Metamaterials

- Traditional instruments have restricted sound possibilities compared to electronic music
- Extended techniques have been used since the 60th to enlarge sound properties
- Many new instruments have been suggeted (composite materials, sensor techniques, hybrid instruments combining acoustics and electronics)

Metamaterials is another way to built new sounds

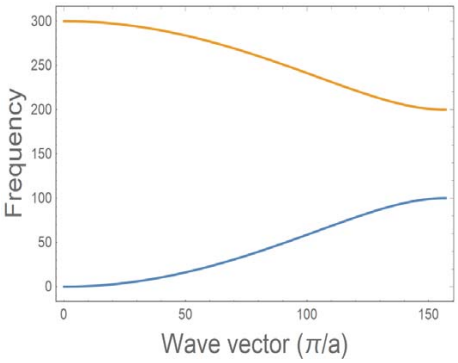
- Negative Young‘s modulus, negative density, negative refraction index
- Cloaking of waves
- Band gaps in spectrum
- Dispersion relations compressing or stretching harmonic overtone structures
- Viscoelastic damping of single frequencies

String with additional masses

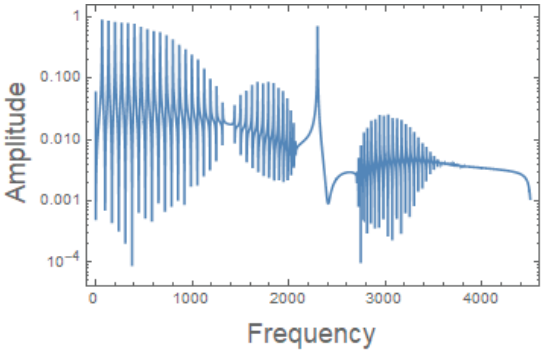


Dispersion relation

Theory of phononic crystal



Finite-Difference Physical Model

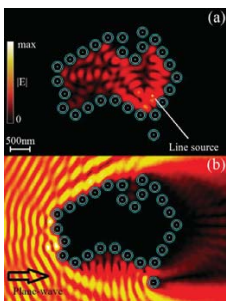


Frame Drum with cirle

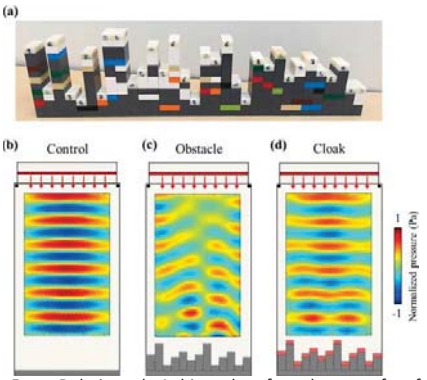
- Regular frame drum
- Circle of magnets added

⇒ Cloaking behaviour of waves
⇒ Frequency bands enhance / reduced

⇒ New articulatory possibilities

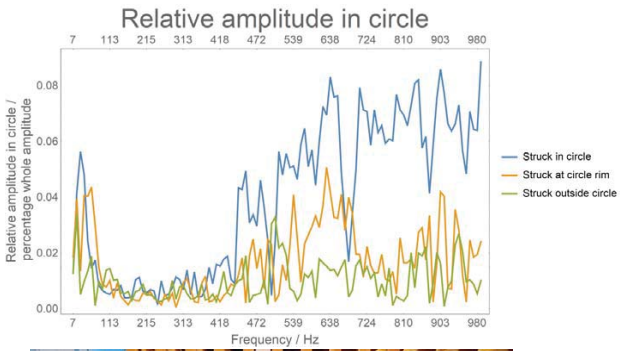


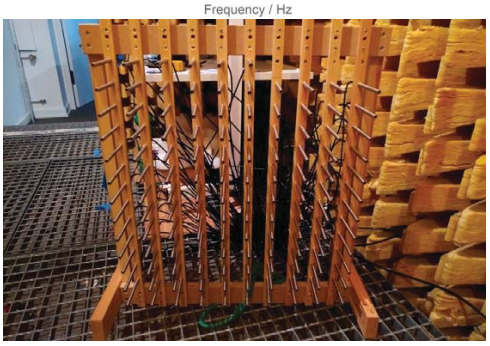
From: Mirzaei et al.: Optical Metacave. Phy. Rev. Lett. 115, 2015.



From: Dubois et al.: A thin and conformal metasurface for illusion acoustics of rapidly changing profiles. Appl. Phy. Lett. 110, 2017.

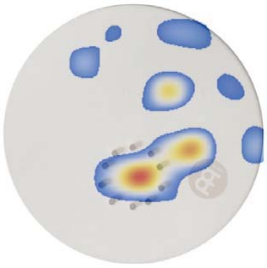
Commulation of mode maximum amplitudes on the drum for 34 Hz – 1kHz for three strike positions:



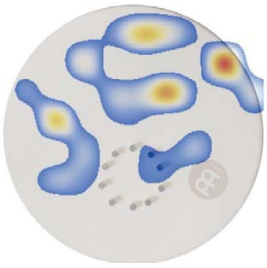


Striking in circle

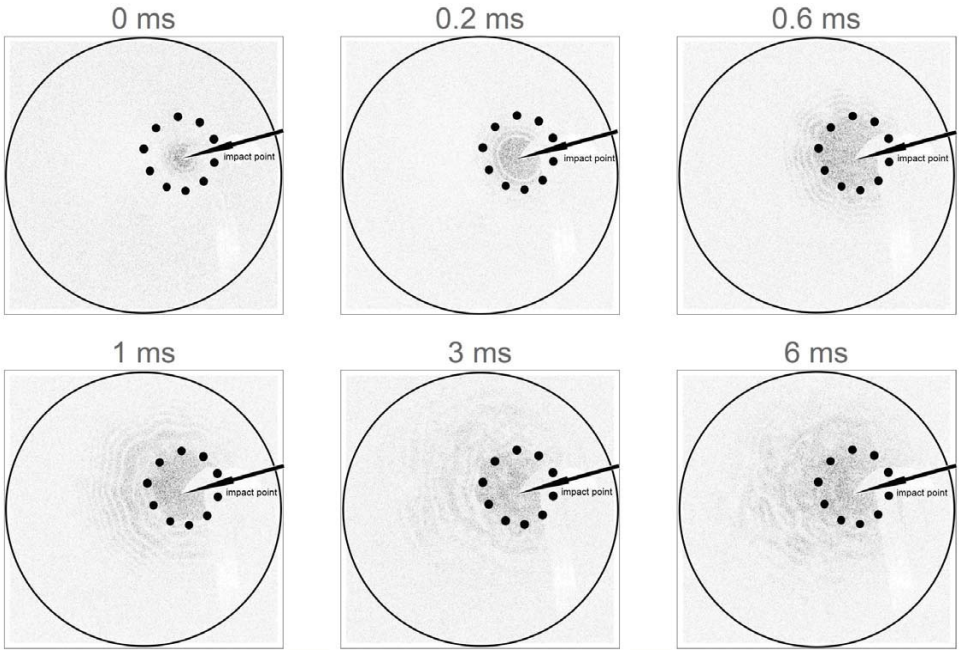
Striking at circle rim



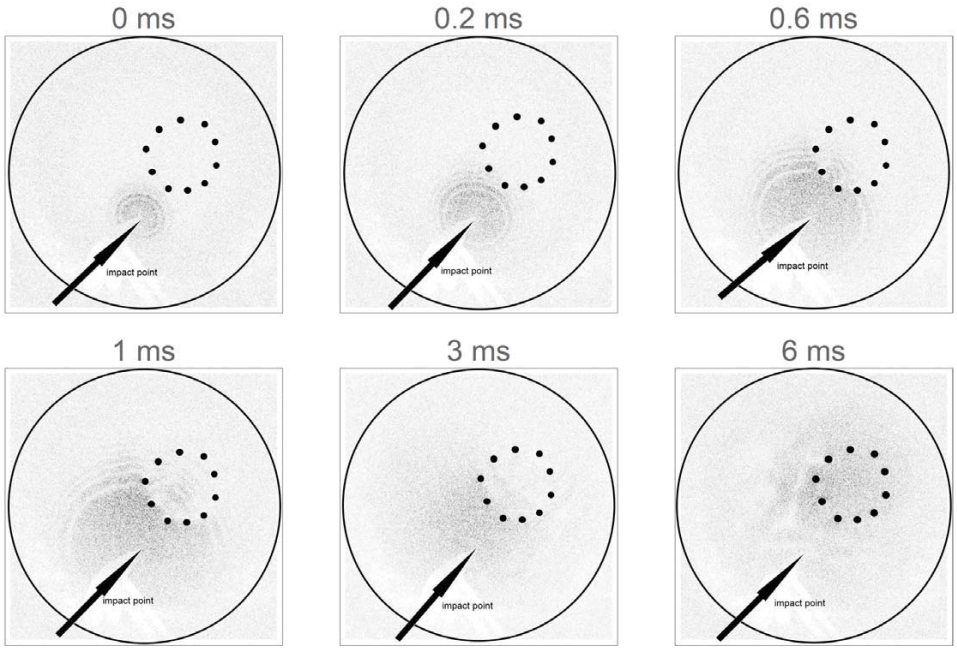
Striking outside circle



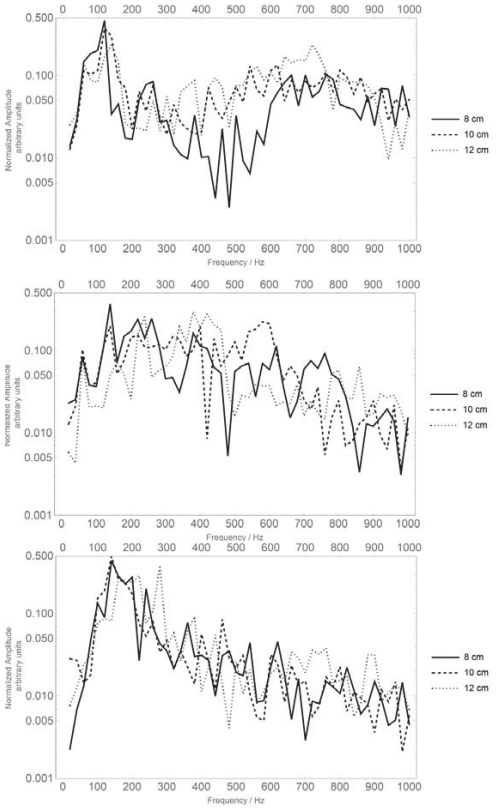
Laser interferometry, transient recording 10 000 fps
Striking in the circle



Laser interferometry, transient recording 10 000 fps
Striking outside the circle



Frequency band-gap



Session 4:
Social Network and Data Analysis
(Chair: Ryozo Kiyohara)

Linking Anonymous SNS Accounts and Real-World People through Profiling

Eina Hashimoto[*], Masatsugu Ichino[**] , and Hiroshi Yoshiura[*]

[*,**]Graduate School of Informatics and Engineering, University of Electro-Communications, Japan
[*]{eina.hashimoto, yoshiura}@uec.ac.jp, [**]ichino@inf.uec.ac.jp

Abstract -Social network services (SNSs) are widely used to enrich communication. An important feature of many SNSs is that they enable accounts and posts to be anonymous. Anonymity has benefits and drawbacks. It enhances privacy, but if it is compromised, people who rely on anonymity can be negatively affected. Anonymity encourages people cause malicious behaviors such as revealing confidential information and posting offensive contents. Clarifying the limitation of SNS anonymity is therefore important to warn people of the privacy risks and to deter them from the illegal behaviors.
Related work on clarifying the limitations of SNS anonymity has focused on identifying accounts on different SNSs that belong to the same person. The work reported here focused on linking SNS accounts to real people, i.e. to data that directly represent them, such as the information in resumes. The linking is bidirectional: given an anonymous SNS account, the owner can be identified, and given a person's name, the person's anonymous SNS account can be identified. The method infers the profile of the unknown owner of an SNS account and compares the inferred profile with the information in each available resume. Machine learning is used to estimate the probability of the account owner having the attributes described in each resume. The proposed method was evaluated using the resumes and anonymized Twitter accounts of 78 volunteers, showing that anonymity may be compromised even when a small number of SNS posts are available or when SNS accounts are hidden in a large number of noise accounts.

Keywords: Privacy, Security, De-anonymization, Profiling, SNS

1 INTRODUCTION

Social network services (SNSs) are widely used to enrich communication. They are used not only for communication among friends and family members but also for job hunting, marketing, branding, and political communication such as among political activists. An important feature of many SNSs is that they enable accounts and posts to be anonymous. Anonymity has benefits and drawbacks. It enhances privacy, which promotes free expression of opinions. However, if it is compromised, people who rely on anonymity can be negatively affected such as by being criticized or even by being intimidated. Anonymity can also encourage people to behave maliciously such as by revealing organization's confidential information or posting offensive contents. Clarifying the limitations of SNS anonymity is therefore important for making people aware of the privacy risks and deterring bad actors from behaving maliciously.

The limitations of SNS anonymity have been actively studied in the areas of de-anonymization and re-identification [1–10]. Methods previously studied, however, focused on identifying accounts on different SNSs that belong to the same person. They are thus indirect because knowing that account-1 and account-2 are used by the same person does not directly reveal the person's identity.

In contrast, the work reported here focused on linking SNS accounts to real people, i.e. to data that directly represent them, such as the information in resumes. The linking is bidirectional: given an anonymous SNS account, the owner's resume can be identified, and given a person's resume, the person's anonymous SNS account can be identified.

While an SNS account has posts of natural-language texts and pictures, a resume is basically a list of keywords (e.g. names of schools from which the person graduated). Because an SNS and a resume contain different kinds of data, they cannot be linked straightforwardly. Our proposed method infers the profile of the unknown owner of an SNS account and compares the inferred profile with the profile described in each available resume. Machine learning is used to estimate the probability of the account owner having the attributes described in each resume. The method was implemented and evaluated using the resumes and anonymized Twitter accounts of 78 volunteers to determine its effectiveness.

Our contributions are summarized as follows.

- Our proposed method links SNS accounts to resumes that directly and uniquely represent real-world people. It therefore shows the limitations of SNS anonymity more clearly than previous methods.
- The method is novel in using profiling to match different kinds of data, i.e. data in SNS posts and in resumes.
- Evaluation of the method demonstrated that anonymity can be compromised even when only a small number of SNS posts (e.g. 63 posts per account) are available or when SNS accounts are hidden within a large number of noise accounts (e.g. 100,000 accounts).

Section 2 of this paper analyses related work. Section 3 presents an overview of our research. Section 4 describes the design and implementation of our linking method. Section 5 describes the dataset used in our evaluation, and Section 6 describes the evaluation. Section 7 summarizes our achievements and mentions future work.

2 RELATED WORK

2.1 De-anonymization of SNS accounts

Methods for de-anonymization of SNS accounts can be classified into two types, those based on relationships between users and those based on the features of each SNS account. Early methods were relationship based. In 2007, Backstrom et al. reported that they had de-anonymized anonymous SNS accounts on a social network by searching the network for subgraphs of known human relationships and identifying the subgraph nodes that represented users and friends [1]. In 2009, Narayanan et al. reported a method based on subgraph matching for linking Twitter accounts to the Flicker accounts of the same users [2]. Though interest in graph-matching-based methods waned in the early 2010s, it has recently resurged. In 2014, Niliadeh et at. reported an improved subgraph-matching algorithm that maintains precision even when fake nodes and edges are incorporated into the graphs of the two SNSs for which accounts are to be linked. They first divided the two target graphs into subgraphs representing communities, then matched the communities in the two corresponding subgraphs, and then matched the nodes and edges [3]. Gulyas et al. improved the algorithm's precision by incorporating a new similarity metric [4][5]. Lee et al. further improved the accuracy of subgraph matching by using machine learning to tune the similarity metrics [6].

The feature-based methods include the one reported in 2012 by Goga et al. Their method identifies accounts in different social networks (Yelp, Twitter, Flickr, and Twitter) belonging to the same person by analysing the features of the posts (geo-location, timestamp, writing style, etc.) [7]. Heuristics are used to quantify the similarity between accounts in terms of each feature, and then a machine-learning algorithm using binary logistic regression is used to quantify total similarity. In the same year, Narayanan et al. reported the identification of blogs posted by the same person. The identification was done by using machine-learning algorithms including SVM and linear discriminant analysis to quantify writing style similarity [8]. In 2014, Almishari et al. reported the identification of Twitter accounts used by the same person by using a naïve Bayes classifier to quantify writing style similarity [9]. In 2016, Overdorf et al. reported the identification of blogs, Twitter feeds, and Reddit comments posted by the same person. They overcame the difficulty of cross-domain linking (such as large differences in sentence length) by using an ensemble of SVM and logistic regression [10].

2.2 Problems with Previous Methods

Both types of methods (relationship-based and feature-based) link accounts and posts belonging to the same person. They thus do not work when the target person has only one SNS account. They do work when the person has two accounts either on one SNS or on two SNSs, but if both accounts are anonymous, they can only increase the amount of information about the target person. They cannot identify the person.

The relationship-based methods do not work when the number of targeted persons is small, meaning that the targeted persons are represented by a small graph with few nodes that can be matched to many parts of the other graph. The feature-based methods use writing style as the main feature. They use a machine-learning algorithm to learn the writing style in posts on an SNS account and then check posts on other accounts to determine whether they have a similar writing style. This means that these methods work only for text data.

3 LINKING SNS ACCOUNTS AND PEOPLE

3.1 Using Resumes

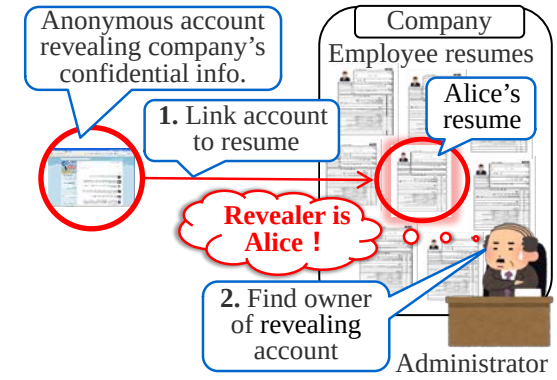
Previous methods link an SNS account or post to another SNS account or post belonging to the same person and are thus insufficient to show the limitations of SNS anonymity. Our approach is to link an SNS account to a person in the real world. Because a computer cannot actually deal with a real person, we link an SNS account to data directly representing a person. We use resumes as the data source. A resume directly represents a person. Most organizations, such as companies and public institutes, have resumes for their members. Schools (elementary through university) have student data records that are similar to resumes. "Resumes" are thus readily available.

An SNS account has posted text and picture data, account owner profile data, and meta-data. Although a resume may include a picture of the person, a picture of the account owner is not usually posted on an SNS account. It is thus difficult to link an SNS account and resume on the basis of pictures. The profile data and meta-data for an SNS account do not usually include information found in a resume. We therefore use text posted on an SNS account and compare it to the resume.

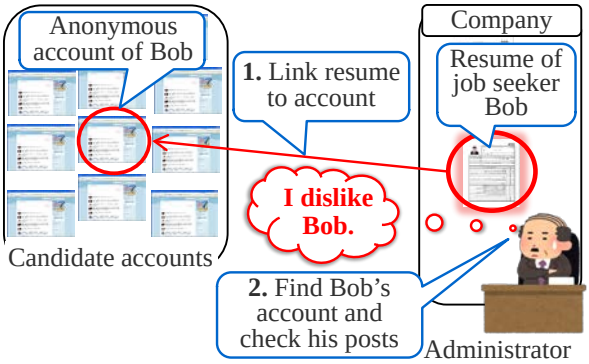
3.2 Bidirectional Linking

Our method links SNS accounts and resumes bidirectionally: given an anonymous SNS account, it identifies the resume of the account owner and, given a resume, it identifies the SNS account of the person described in the resume. Figure 1 (a) shows an example of linking an SNS account to a resume. In this example, we assume that an employee, Alice, revealed confidential company information on an anonymous account. The company uses our method to link the anonymous account to the resume of the revealer, thus identifying the revealer. From among all the resumes in the company's files, the method determined that Alice's resume is the one most likely to describe the person to whom the SNS account belongs.

Figure 1 (b) shows an example of linking a resume to an SNS account. In this example, we assume that a job seeker, Bob, submitted a resume to the company. The company uses our method to link Bob's resume to one of several candidate SNS accounts, enabling it to examine Bob's posts. From among all the candidate SNS accounts, the method



(a) Application to security by linking account to resume



(b) Application to privacy by linking resume to account

Figure 1: Examples of bidirectional linking.

identified the account most likely used by the person described in the resume.

3.3 Application to Security and Privacy

Our method can be used for both security and privacy purposes. For security, as shown in Figure 1 (a), it can be used to identify a person who maliciously revealed confidential information, thus enabling the company to punish the revealer. The company can also use it to deter people from such revelations. Similarly, it can be used to deter people from posting copyrighted, bullying, or offensive content. On the other hand, our method can be used for maintaining privacy as well. As shown in Fig. 1 (b), it can be used to clarify the risk of a job seeker's privacy being compromised, thus warning him or her of the privacy risk of using SNS anonymously. Similarly, it can clarify the risk of a candidate for promotion and an opponent of an organization of their anonymous SNS accounts being identified.

4 LINKING METHOD

4.1 Technical Challenges

Linking SNS accounts and resumes involves three technical challenges. The first is linking different kinds of data—texts posted on an SNS account consist of natural-language sentences while a resume is basically a list of keywords representing the attributes of the person (such as names of schools from which the person graduated). Second, it is not easy to obtain resumes because resumes contain sensitive information. While it may be easy to obtain the resume of the target person, obtaining the resumes of other people may be difficult. Analysing and modelling the resume of the target person is difficult without using other resumes. Third, the number of posts differs greatly among accounts; thus, only a small number of posts may be available for some candidates.

Because a resume does not include links to other people's resumes, and there may be only a few resumes, we adopted feature-based linking rather than relationship-based linking. We use machine learning, which was recently applied to feature-based linking. We first considered building a model of a pair of SNS accounts and a resume, all belonging to the same person. Such a model could be used to determine whether posts on an anonymous SNS account and information in a resume belong to the same person. However, the machine learning for such building needs training data collected from SNS posts and resumes that belong to the same person, and collecting such training data is difficult.

Next, we considered building a model of each resume that might be used to determine whether posts on an anonymous SNS account belong to the person described in the resume. However, the machine learning for this building needs training data collected not only from the resume of the target person but also from the resumes of other people as negative samples, and collecting such negative samples is difficult. Moreover, it is doubtful that a model learned from resumes could be successfully applied to SNS posts, which are quite different from the text in resumes.

We also considered building a model of each anonymous SNS account that might be used to determine whether a resume described an account's unknown owner. However, it is doubtful that a model learned from SNS posts could be successfully applied to resumes, which are quite different from SNS posts.

4.2 Linking by Profiling

To overcome the challenges mentioned above, we devised a method of linking based on profiling. Profiling the owner of an SNS account has been well studied, with machine learning being used to infer the attributes of the SNS owner, such as age, gender, and marital status [11], medical condition [12], and political affiliation, ethnicity, and coffee brand preference [13].

Figure 2 illustrates the basic idea of our linking method. A resume describes the profile of a person. We therefore infer the profile of the owner of an SNS account and compare the inferred profile with that described in the resume to determine whether the owner of the SNS account is the person described in the resume.

Figure 3 illustrates how our method works. It first extracts attributes and their values from the resume (gender=female, age=30s, hobby=dancing, and so on). Next, profiling is performed on the anonymous SNS account to estimate the probability of the account owner having the same attribute values (being female with a probability of 82%, being in 30s with a probability of 54%, loves dancing with a probability

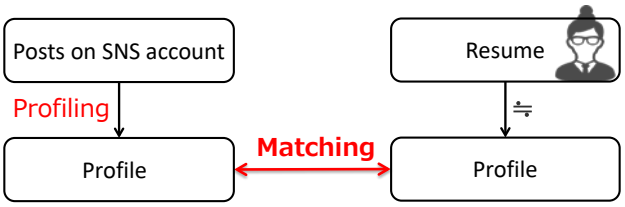


Figure 2: Basic idea of proposed method.

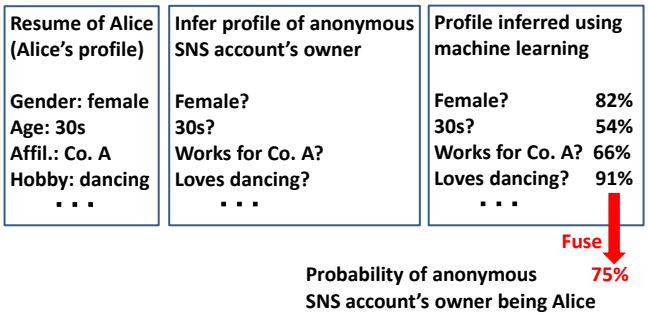


Figure 3: Basic processes of proposed method.

of 91%, and so on). The probability is estimated by using machine learning, as detailed in Section 4.3. Finally, the probability estimated for each attribute is fused to quantify the probability of the anonymous owner of the SNS account being the person described in the resume. This process is done for each pair of SNS account and resume. On the basis of the probability of each pair belonging to the same person, each SNS account is linked to the resume that is most likely to belong to the same person.

4.3 Using Machine Learning

Figure 4 illustrates how machine learning is used to identify the person that is most likely to be the anonymous owner of an SNS account. We assume that we have resumes of the candidate SNS owners (Alice, Bob, and so on). If, for example, Alice's resume states that Alice is female, in her 30s, with dancing as her hobby, machine learning is used to build models of these attribute values. The model for Alice is composed of these models. A model for another person, e.g. Bob, is built similarly.

When we want to de-anonymize the anonymous owner X of an SNS account, we input posts from this account into the model of each person. When the posts from an unknown account are input into, for example, Alice's model, they are input into the model of each attribute values (models for woman, 30s, dance lover, and so on). The model of each attribute value outputs the probability of person X having the corresponding attribute value (e.g. the probability of X being female). This process is profiling of person X along with Alice's resume. The probability output from the model of each attribute value is fused into the probability of person X being Alice. Similarly, the probability of person X being Bob, Carol, and so on is estimated. The method then selects the person with the highest probability as the person who is most likely to be person X.

A model of an attribute value (e.g. model for woman) is built by machine learning with training data consisting of positive and negative samples. These samples are posts corresponding to the attribute value, e.g. posts of women

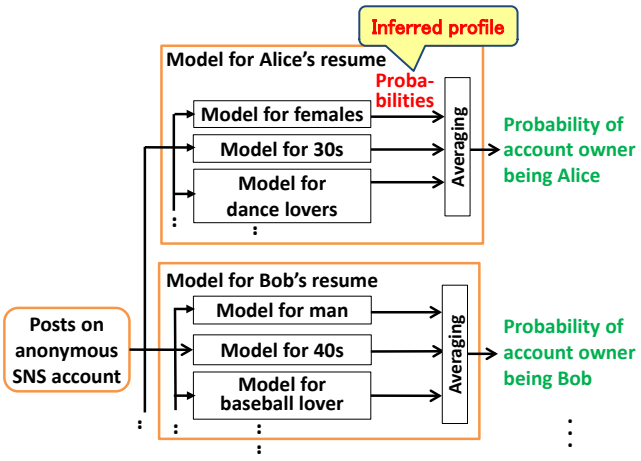


Figure 4: Inferring profile of SNS account owner using machine learning.

and men. They are collected from public SNS accounts that are randomly selected.

The problem of linking different types of data is thus solved by matching profiles inferred from the data. Moreover, the problem of a lack of negative training data (i.e. resumes of persons other than the target person) is solved by building models of attribute values described in the resume of the target person rather than building a model of the resume itself. Because these attribute values (gender, age, hobby, etc.) are general, we can collect training data from public SNS accounts.

4.4 Implementation

We considered two features of posts on SNS accounts: the number of times each word appears in the posts (the "bag-of-words" feature) and the existence of each word in posts (the "binary" feature). We also considered several machine-learning algorithms, linear SVM, SVM with a Gaussian kernel, naïve Bayes, logistic regression, random forest, and XGBoost. We selected the bag-of-words/XGBoost combination as it had shown the best performance in a preliminary evaluation. We considered the use of different types of words to build the bag-of-words: (1) nouns, (2) all words except stop words, (3) all words. Since "all words" had shown the best performance in a preliminary evaluation, we used all words in the bag-of-words. We extracted the words from posts by using the Mecab tool for natural language processing [14].

We considered two methods for fusing the probability output from each attribute model to obtain the probability of the owner of the SNS account being the person described in the resume: (1) averaging and (2) using meta-learning to optimally combine the probability values. Though meta learning might have provided better precision, it would have needed extra training data, i.e. we would have needed not only training data for attribute-value models but also those for training the meta model. The training data for the meta model must have been pairs of SNS account and resume belonging to the same person but such data would have been difficult to collect as mentioned in Section 4.1. We therefore used simple averaging.

5 DATA DESCRIPTION

5.1 Sample Data from Volunteers

Hereafter we abbreviate "social network account" as "account". We obtained Twitter accounts and resumes from 78 volunteers (27 students at our university and 51 other people (both students and working people). The tweets and resumes were originally written in Japanese and were translated into English when necessary. Figure 5 shows their demographics. We collected at most the 3000 latest posts from each of their accounts during May 2018.

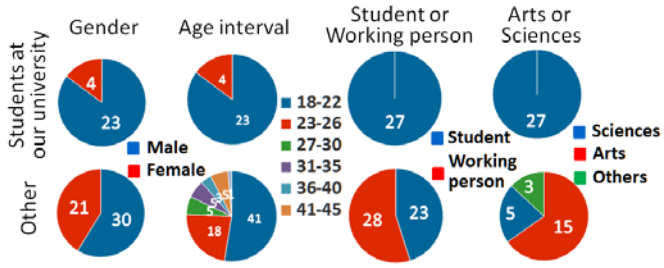


Figure 5: Volunteer demographics.

The attributes in their resumes were determined by referring to the Japanese standard for the resumes of students seeking employment. We also included the attribute "type of job" for the working volunteers. We thus used ten attributes: (1) gender, (2) age interval, (3) current address, (4) hometown address, (5) educational history, (6) type of job, (7) favourite subjects, (8) strong points, (9) hobbies and clubs, and (10) qualifications. We used age interval instead of birthdate because birthdate information is sensitive. Since educational history is generally complex, we simply used the university that the volunteers were attending or from which they had graduated as representative information. Some strong points, such as endurance and friendliness, are difficult to model. We therefore recategorized some strong points such as mathematics as a favourite subject or as a hobby and ignored the remaining ones.

Not all volunteers provided information for all attributes. Some did not have some attributes (e.g. most students did not have a job), and some were reluctant to provide information for some attributes. Some volunteers had multiple values for favourite subjects, hobbies and clubs, and qualifications. For example, one volunteer had only one hobby (dancing) while another had three (traveling, playing football, and playing board games). The number of attribute values described in a resume thus depended on the volunteer. The number was 11.14 on average, with a minimum of 6, a maximum of 28, and a standard deviation of 3.7.

Linking the accounts and resumes of the 27 students at our university was expected to be more difficult than linking those of the other 51 volunteers because the 27 students are quite similar to each other demographically. They were in one of two departments (informatics or electronics), which are in neighbouring buildings. Their current addresses are close to the university and close to each other. Because the Informatics and Electronics Departments share many subjects such as computer architecture, programming, and signal processing, they had similar favourite subjects. Their daily schedules were also similar. Thus, their SNS posts and resumes were similar.

5.2 Training Data from Public Accounts

As mentioned in Section 4.3, the model for each attribute value was built using positive and negative examples, i.e. posts from SNS accounts that did and did not have the attribute value. Collecting positive examples for training data was done automatically using the TwiPro tool [15], which searches the Internet for social network accounts for which the user profile includes a given attribute value (e.g. gender = female). While this search works for most attribute values, it cannot collect a sufficient number of examples for unusual attribute values such as "qualification = license for piloting small ships". Collecting negative examples for training data was easier—the same tool was used to search for social network accounts for which the user profile did not include a given attribute value. We used accounts having more than 1000 posts and collected up to 3000 posts on each account. For each attribute value, we used 100 accounts for collecting positive examples and also 100 accounts for collecting negative example posts, except for the unusual attribute values. We built models of these unusual attribute values when examples from more than 10 accounts were available.

6 EVALUATION

6.1 Linking Accounts to Resumes

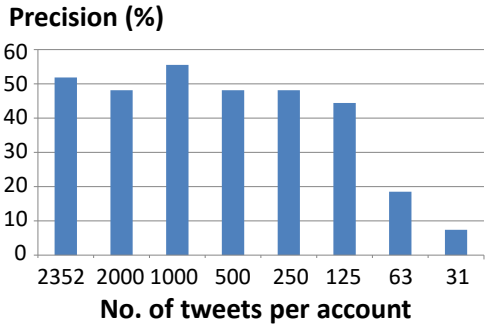
In the first experiment, our method linked each of the 78 accounts to the resume describing the person most likely to be the same person as the anonymous owner of the account. As shown in Table 1, 39 accounts were correctly linked to the resume; i.e. for each of these accounts, the correct resume had the top score among the 78 resumes where the score is the probability calculated for each pair of account and resume as mentioned in Section 4.3. For 67 accounts, the correct resume was among the top 10%, and, for 72 accounts, the correct resume was among the top 20%. Table 1 also shows the results broken down by the 27 students at our university and the 51 other volunteers.

Table 1: Number of accounts correctly linked to resumes

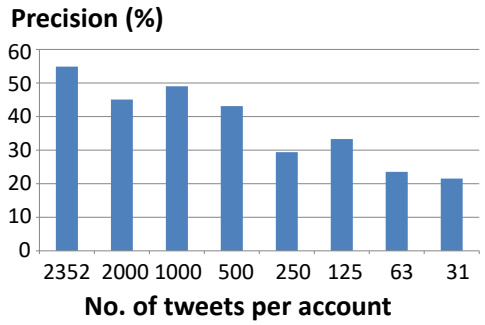
	All 78 volunteers	**27 students at our university**	**51 other volunteers**
Top	39 (50%)	14 (51.8%)	28 (54.9%)
Top 10%	67 (85.8%)	21 (77.7%)	44 (86.2%)
Top 20%	72 (92.3%)	26 (96.2%)	46 (90.1%)

We analysed the features of the 39 volunteers whose accounts were correctly linked to their resumes. The number of attribute values described in their resumes was 11.41 on average, with a minimum of 6 and a maximum of 28. As mentioned in Section 5.1, the number for all resumes was 11.14 on average, with a minimum of 6 and a maximum of 28. The number of attributes described in the resume thus did not affect identifiability.

The 39 volunteers were classified into two types. The first type included those volunteers with at least one attribute



(a) Precision for 27 students at our university



(b) Precision for 51 other volunteers
Figure 6: Precision with less data.

value for which he or she was scored the highest. For example, one such person had "hometown = City A in Kumamoto Prefecture". The model "hometown = City A in Kumamoto Prefecture" judged that the probability of the target person's hometown being City A in Kumamoto Prefecture was 0.996 while the probability of any other person's hometown being City A in Kumamoto Prefecture was less than 0.05. For another example, the model "hobby = archery" judged that the probability of the target person's hobby being archery was 0.979 while the probability of any other person's hobby being archery was less than 0.65.

The second type included those volunteers with multiple attributes for which he or she was scored high. For example, the models "hometown = City B in Saitama Pref.", "hobby = dancing", and "hobby = playing keyboard" judged that the probability of one person having each of these attribute values was more than 0.9.

Figure 6 shows the number of accounts that were correctly linked to a resume when we used less than all the tweets posted on each account. The number did not drop much compared with using all the tweets (average of 2352 tweets per account). Five out of 27 accounts (19%) of the students at our university and 12 out of 51 (24%) of the other volunteers were correctly linked even when we used only the latest 63 tweets. These results show that anonymity may be compromised even when a small number of posts are available.

From Table 1, we can also see that the students at our university were more difficult to identify than the other volunteers. This is because they were similar to each other, as described in Section 5.1.

6.2 Linking Resumes to Accounts

In the second experiment, our method linked each of the 78 resumes to the account for which the owner was most likely to be the person described in the resume. As shown in Table 2, 16 resumes were correctly linked; i.e. for each resume, the correct account had the top score among the 78 accounts. For all resumes, the scores of the correct accounts were in the top 10%. We also attempted to link the 78 volunteers to their accounts after their accounts were shuffled with 1000 randomly collected accounts (noise accounts). As shown in Table 2, 9 of the 78 resumes (12%) were correctly linked. The correct accounts for 59 resumes (76%) were in the 100 highest scoring accounts. The correct accounts for 12 resumes (15%) were in the 100 highest scoring accounts even when they were shuffled with 100,000 noise accounts. Once the correct account had been included in 100 candidates, we could have manually scrutinized these candidates to identify the correct one. These results show that the anonymity of SNS accounts may be compromised even if they are hidden among a large number of noise accounts.

Table 2: Number of resumes correctly linked to accounts when noise accounts were added

No. of noise accounts	0	1000	10000	100000	300000
Top	16 (21%)	9 (12%)	5 (6%)	4 (5%)	1 (1%)
Within 100 highest	78 (100%)	59 (76%)	20 (26%)	12 (15%)	7 (9%)

7 CONCLUSION

We have developed a method for de-anonymizing SNS accounts to clarify the limitations of SNS anonymity. While previous methods linked accounts in different SNSs belonging to the same person, our proposed method links an SNS account to the resume of the account owner. Because a resume directly and uniquely represents a real-world person, our method shows the limitations of SNS anonymity more clearly than previous methods. Because most organizations have resumes for their members, our method is easy to implement.

To link different types of data, i.e. SNS accounts and resumes, our method infers the profile of the unknown owner of each SNS account. It then compares the inferred profile with that described in each resume to identify the best-matched resume for each SNS account. Because our profiling uses models of attribute values rather than models of resumes, training data can be collected from public SNS accounts. Evaluation of the proposed method showed that anonymity may be compromised even when a small number of SNS posts are available or when SNS accounts are hidden in a large number of noise accounts.

Future work includes improving the method for fusing the model attribute values by using meta-learning rather than simple averaging. It also includes extending the method to cope with cases in which only some of the resume information is available rather than all the information.

REFERENCES

[1] R. Backstrom, C. Dwork, and J. Kleinberg, Wherefore Art Thou R3579X? Anonymized Social Networks, Hidden Patterns, and Structural Steganography, 16th International World Wide Web Conference, pp. 181-190 (2007).

[2] A. Narayanan and V. Shmatikov, De-anonymizing Social Networks, 30th Symposium on Security and Privacy, pp.173-187 (2009).

[3] S. Nilizadeh, A. Kapadia, and Y. Ahn, Community-Enhanced De-anonymization of Online Social Networks, 21th ACM Conference on Computer and Communications Security, pp.537-548 (2014).

[4] G. Gulyas and S. Imre, Measuring Importance of Seeding for Structural De-anonymization Attacks in Social Networks, Pervasive Computing and Communications Workshops (In conjunction with 12th IEEE International Conference on Pervasive Computing and Communications), pp.0-6 (2014).

[5] G. Gulyas, B. Simon, and S. Imre, An Efficient and Robust Social Network De-anonymization Attack, 15th Workshop on Privacy in the Electronic Society, pp.1-11 (2016).

[6] W. Lee, C. Liu, S. Ji, P. Mittal, and R. Lee, Blind De-anonymization Attacks using Social Networks, 16th Workshop on Privacy in the Electronic Society, pp.1-4 (2017).

[7] O. Goga, H. Lei, S. Krishnan, G. Friedland, R. Sommer, and R. Teixeira, On Exploiting Innocuous User Activity for Correlating Accounts Across Social Network Sites, ICSI Technical Reports - University of Berkeley (2012).

[8] A. Narayanan, H. Paskov, et al., On the Feasibility of Internet-scale Author Identification, 33rd IEEE Symposium on Security and Privacy, pp.300–314 (2012).

[9] M. Almishari, M. Kaafar, E. Oguz, and G. Tsudik, Stylometric Linkability of Tweets, 13th Workshop on Privacy in the Electronic Society, pp.205-208 (2014).

[10] R. Overdorf and R. Greenstadt, Blogs, Twitter Feeds, and Reddit Comments: Cross-domain Authorship Attribution, 16th Privacy Enhancing Technologies Symposium (3), pp.155-171 (2016).

[11] G. Kótyuk and L. Buttyan, A Machine Learning based Approach for Predicting Undisclosed Attributes in Social Networks, 4th IEEE International Workshop on Security and Social Networking, pp.361–366 (2012).

[12] H. Mao, X. Shuai, and A. Kapadia, Loose Tweets: An analysis of Privacy Leaks on Twitter, 10th ACM Workshop on Privacy in the Electronic Society, pp.1-12 (2011).

[13] M. Pennacchiotti and A. Popescu, A Machine Learning Approach to Twitter User Classification, 5th International AAAI Conference on Weblogs and Social Media, pp. 281–288 (2011).

[14] T. Kudo, Mecab: Yet Another Part-of-speech and Morphological Analyzer, https://taku910.github.io/mecab/

[15] TwiPro: Searching Profiles of Twitter Users, http://twpro.jp/ (In Japanese)

Examination of Analysis Method Based on Location Information and Time Series Change using Geotagged Tweet for Regional Analysis

Masaki Endo*, Masaharu Hirota**, Makoto Imamura***, and Hiroshi Ishikawa****

* Division of Core Manufacturing, Polytechnic University, Japan
endou@uitec.ac.jp
** Faculty of Informatics, Okayama University of Science, Japan
hirota@mis.ous.ac.jp
*** School of Information and Telecommunication Engineering, Tokai University, Japan
imamura@tsc.u-tokai.ac.jp
**** Graduate School of System Design, Tokyo Metropolitan University, Japan
ishikawa-hiroshi@tmu.ac.jp

Abstract— Because of the popularization of Social Networking Services (SNSs), it is possible to acquire large amounts of data in real-time. For this reason, various studies are being conducted to analyze social media data and to extract real-world events. Among them, a salient advantage of analysis using positional information is that one can accurately extract events from target areas of interest. We proposed a real-time analysis of tourist information by a simple method using a moving average of geotagged tweets on Twitter. However, data having position information in social media data remain few: the data amount might be insufficient for analysis. Therefore, we are assessing a method for real-time analysis using data with location information that has been accumulated over a certain period of time. In this research, we are studying a method of regional analysis by position information and time series change using tweets with Twitter position information. As a result, it may be possible to use it as a complement to the analysis method using moving averages proposed by us. Herein, we explain the results obtained from area analysis using the proposed method.

Keywords: big data; big data utilization technology; location information; time series; Twitter

1 INTRODUCTION

In our everyday life, because of the wide dissemination and rapid performance improvement of various devices such as smartphones and tablets, diverse and vast data are generated on the web. SNSs have become especially popular because users can post data and various messages easily. Twitter [1], an SNS that provides a micro-blogging service, is used as a real-time communication tool. Numerous tweets have been posted daily by vast numbers of users. Twitter is therefore a useful medium to obtain, from a large amount of information posted by many users, real-time information corresponding to the real world.

By analyzing the information sent by these SNSs, the possibility exists of obtaining useful information in real time. We are conducting research related to providing tourist information to travelers. Therefore, this study specifically examines the provision of real-time sightseeing information.

Herein, we describe the provision of information to tourists using web contents. Such information is useful for tourists, but providing timely and topical travel information entails high costs for information providers because they must update the information continually. Today, providing reliable information related to local travel is not only strongly demanded by tourists, but also by local governments, tourism organizations, and travel companies, which bear high costs of providing such information.

For that reason, providing current, useful, real-world information for travelers by ascertaining changes of information according to seasons and time zones of the tourism region is important for the travel industry. It is possible to disseminate information using the popular SNS, but organizations that can actually do the work are limited by human resources and cost. Therefore, analysis using an SNS that can provide useful data leading to real-time information provision is one means of overcoming this difficulty.

To solve this problem, much research to analyze SNS data is currently being conducted. Research using Twitter is one branch of investigation. Because tweets comprise short sentences, a location can be estimated if a tweet includes the place name and the facility name, but if such information is not included, identifying the location from a tweet might be difficult. For this reason, research using tweets with location information or tweets which give location information in the tweet itself is being conducted. Because geotagged tweets can identify places, they are effective for analysis. Nevertheless, few geo-tagged tweets exist among the total information content of tweets. It is therefore not possible to analyze all regions. For that reason, we also use geotagged tweets to conduct research using information interpolation to estimate the position around the area that is not specified by the position information [2].

Currently, we are considering a method for real-time analysis by collecting temporal and spatial information for a certain period of time using only geotagged tweets, which are said to have a small amount of information. This report presents an experimental approach.

The remainder of the paper is organized as follows. Chapter 2 presents earlier research related to this topic. In Chapter 3, we propose a method for real-time analysis using data collected for a certain period. Chapter 4 describes

experimentally obtained results for our proposed method and a discussion of the results. Chapter 5 presents a summary of the contributions and expectations for future work.

2 RELATED WORK

Various studies are being conducted using SNS position information. Omori et al. [3] proposed a method to extract geographical features such as coastlines using tags of photo sharing sites with geotags. Sakaki et al. [4] proposed a method to detect events, such as earthquakes and typhoons based on a study estimating real-time events from Twitter. By analyzing the Twitter text stream, Pratap et al. [5] proposed a solution to optimize traffic control by considering previous traffic analysis methodology and social data in real time. Various analytical methods have been proposed for analyzing SNS using position information and time series information. However, analysis of data in which large amounts of position information and time series information exist is mainly addressed. Few research efforts examine information using only a few data.

Some research has examined visualization. Nakaji et al. [6] proposed using a geotagged and visual feature of a photograph and suggested a way to select photographs related to a given real event from geotagged tweets. They developed a system that can visualize real-world events on online maps. In the GeoNLP Project [7], we are developing a geotagging system that extracts location descriptions such as place names and addresses contained in natural language sentences. The system provides metadata about where the sentences are descriptions. It is also offered as open source software. These studies are very useful for extraction of specific designated events and for analysis of preregistered places. However, another discussion must be held about automatically extracting events and identifying new places.

As described above, conducting research using geotagged tweets for places with small information amounts and new events and places represents a new approach. Therefore, this research was conducted to identify events and places in real time using accumulation of information and differences in space-time space.

3 OUR PROPOSED METHOD

This chapter presents a description of a method for target data collection using our method of real-time analysis with position information and time series information.

3.1 Data collection

Here, we explain the data collection target for this research. Geotagged tweets sent from Twitter are the collection target. The range of geotagged tweets includes the Japanese archipelago (120.0°E ≤ longitude ≤ 154.0°E and 20.0°N ≤ latitude ≤ 47.0°N) as the collection target. Collection of these data was done using a streaming Application Programming Interface (API) [8] provided by Twitter Inc.

Next, we describe the number of collected data. According to a report by Hashimoto et al. [9], among all tweets originating in Japan, only about 0.18% are geotagged tweets: they are rare among all data. However, the collected geotagged tweets number about 70,000, even on weekdays. On some weekend days, more than 100,000 such messages are posted. We use about 423 million geotagged tweets from 2015/2/17 through 2018/12/26. Therefore, we examined 19 million geotagged tweets in Tokyo for these analyses.

3.2 Preprocessing

This chapter presents preprocessing after data collection. Preprocessing includes reverse geocoding and morphological analysis, with database storage for data collected using the process.

Reverse geocoding identified prefectures and municipalities by town name using latitude and longitude information from the individually collected tweets. We use a simple reverse geocoding service [10] available from the National Agriculture and Food Research Organization in this process: e.g., (latitude, longitude) = (35.7384446N, 139.460910W) by reverse geocoding becomes (Tokyo, Kodaira-Shi, Ogawanishi-machi 2-chome). In addition, based on latitude and longitude information of the collected tweets, data are accumulated by the same place. As data accumulate, the data are saved in mesh form as time elapses.

Morphological analysis divides the collected geo-tagged tweet morphemes. We use the “Mecab” morphological analyzer [11]. As an example, “桜は美しいです” (“Cherry blossoms are beautiful.” in English) is divisible into "(桜 / noun), (は / particle), (美しい / adjective), (です / auxiliary verb), (。 / symbol)".

Preprocessing performs the necessary data storage from the result of data collection, reverse geocoding, and morphological analysis processing. Data used for this study are the tweet ID, tweet posting time, tweet text, morpheme analysis result, latitude, and longitude.

3.3 Analysis method

This chapter presents a description of the method of real-time analysis using position information and time series information.

The analytical method we proposed has the following three stages.

1. Extraction of places by fixed point observation
2. Analysis considering the time series based on 1
3. Analysis using co-occurring words of 2

Therein, 1 is an estimate of the location derived from stationary observation. At such spots, even in places with few tweets, one can discover the location through long-term observation. This method does spot extraction by adding geotagged tweets including specific keywords for long periods at every latitude and longitude.

As presented above, 2 is a method of extracting new spots using spot information accumulated over a long period as a baseline, by consideration of the time series and finding differences.

As shown above, 3, time series analysis including co-occurrence words in 2 and for keywords used in 1 is performed using the results of morphological analysis of tweets. It is a method used because differences in latitude,

longitude, and time series alone might be insufficient to extract differences in data.

Through analyses using these proposed methods, we aim to capture real-time changes in specific areas.

4 EXPERIMENTS

This chapter presents a description of a real-time analysis experiment using the method proposed in Chapter 3.

4.1 Dataset

Datasets used for this experiment were collected using streaming API, as described for data collection in 3.1. Data are geo-tagged tweets from Tokyo during 2015/2/17 – 2018/12/26. The data include about 19 million items. We use these datasets for experiments to conduct the three methods proposed in Chapter 3.

4.2 Experimental method

In this chapter, experiments using the proposed method shown in Chapter 3 are described in 4.2.1. to 4.2.3..

4.2.1. Extraction of places by fixed point observation

This experiment was conducted for Takao-machi, Hachioji, Tokyo: an area of about 4 km east–west and about 2.5 km north–south, as shown in Figure 1. Experimentally obtained results described later are included within the thick frame depicted in Figure 1. For this area, we conducted an extraction experiment with the target word as "cherry blossom" in Japanese as "桜", "さくら", or "サクラ". In all, 65 tweets were found to include a target word.

Figure 1: Target of Takao-machi, Hachioji City.

4.2.2. Analysis in considering of time series based on 1

This experiment was conducted for 4-chome, Myojin-cho, Hachioji city, Tokyo: an area of about 700 m east–west, and about 600 m north–south, as shown in Figure 2. Experimentally obtained results described later are included within the thick frame in Figure 2. For this area, we conduct an extraction experiment with the target word as "ramen" in Japanese as "ラーメン", "らーめん", or "拉麺". In all, 301 tweets were found to include a target word.

Figure 2: Target of Myojin-cho 4-chome, Hachioji City.

4.2.3. Analysis using co-occurring words of 2

This experiment was conducted for Marunouchi 1-chome, Chiyoda-ku, Tokyo: an area of about 1 km east–west and about 1 km north–south, as shown in Figure 3. Experimentally obtained results described later are included within the thick frame in Figure 3. For this area, we set the target word as "ramen", as in the second experiment, in Japanese as "ラーメン", "らーめん", or "拉麺". In all, 6,979 tweets included a target word. As words co- occurring after the object, we target tweets including "𠮷蔵" and "玉", which represents the ramen shop name; the respective numbers of tweets were 273 and 31.

Figure 3: Target area of Marunouchi 1-chome, Chiyoda-ku.

4.3 Experimental result

In this chapter, the results of 4.2.1–4.2.3 experiments explained in the previous chapter are presented.

4.3.1. Extraction of places by fixed point observation

The distributions of geotagged tweets in Takao-machi, Hachioji City including cherry blossoms obtained in the experiment are shown in Figure 4 in 2017 and in Figure 5 in 2018. The interior are of the bold frame in Figure 1 is shown in the table. It is about 265 m measured east–west and about 85 m measured north–south. This cell is obtained by dividing the maximum value and the minimum value of latitude and longitude into 25 and 16 respectively for geotagging tweets. In the current experiment, the number of divisions is specified manually because the target range is different for each experiment. The closer the color of the cell is to black, the more data are shown.

Data extracted for this experiment were very few: 65 for the entire collection period. However, in 2017 and 2018, we confirmed tweets to JR Takao Station, Takao Yamaguchi Station, Takao Station of Ropeway, and Takao Mountain. The correlation coefficient between the extracted spots in 2017 and 2018 was 0.769: high positive correlation was found. The correlation coefficient is calculated by equation (1).

$$\mathrm{Correl}(X-Y)=\frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2\sum(Y-\bar{Y})^2}} \quad (1)$$

$$\mathrm{X}: Number\ of\ Tweet\ of\ 2017_{(latitude,longitude)}$$

$$\mathrm{Y}: Number\ of\ Tweet\ of\ 2018_{(latitude,longitude)}$$

Extraction of spots is possible even with few data. Moreover, various spots can be extracted when using longer periods. Therefore, fixed point extraction of sightseeing spots is regarded as possible through continuing observation of geotagged tweets.

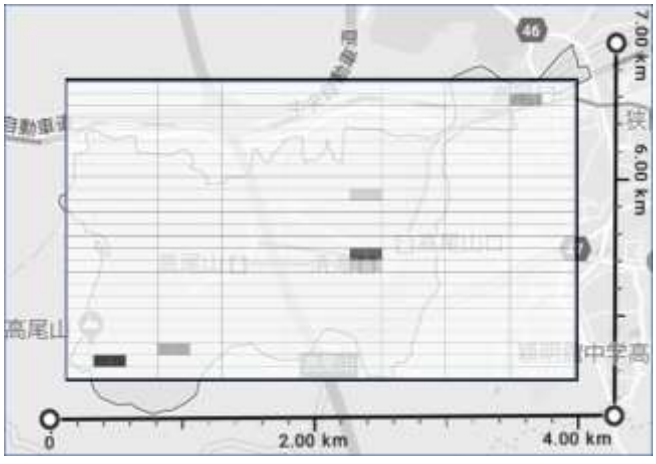


Figure 4: Number of Tweets including Target Words in Takao-machi in 2017.

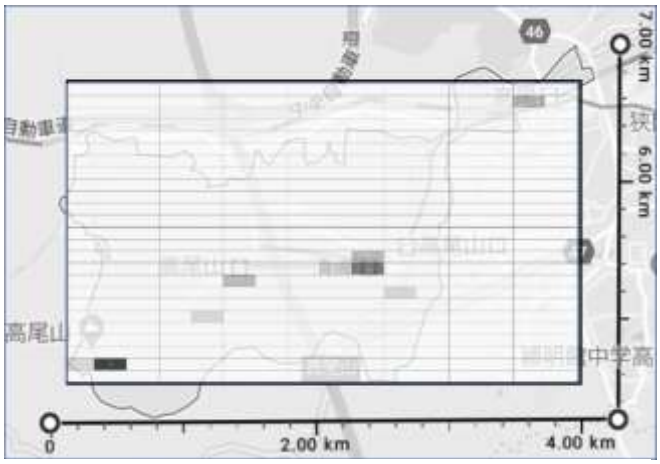


Figure 5: Number of Tweets including Target Words in Takao-machi in 2018.

4.3.2. Analysis in considering of time series based on 1

The results for distribution of the geotagged tweets of Myojin-cho 4-chome, Hachioji City including the ramen obtained in the experiment are shown in Figure 6 in 2016, Figure 7 in 2017, and Figure 8 in 2018. The area inside of the bold frame in Figure 2 is shown in the table. It is about 102 m measured east–west and about 39 m north–south. This cell is obtained by dividing the maximum value and the minimum value of latitude and longitude into 14 and 6 respectively for geotagging tweets.

In this experiment, 301 data were extracted in all collection periods. The target area has three ramen shops, which can be extracted from the table of each year. Furthermore, in 2018, a new ramen shop opened at one point; in 2018 we extracted a new spot (latitude, longitude) = (35.69540009399, 139.345001221). For this reason, the correlation coefficient between data of 2016 and 2017 was 0.991, a high positive correlation was obtained. Nevertheless, the correlation coefficient between 2017 and 2018 was only a weak positive correlation of 0.161. These results demonstrated the possibility of extracting new spots by fixed point observation. These correlation coefficients are calculated using equation (1).

Figure 6: Number of Tweets including Target Words in Myojin-cho in 2016.

Figure 7: Number of Tweets including Target Words in Myojin-cho in 2017.

Figure 8: Number of Tweets including Target Words in Myojin-cho in 2018.

4.3.3. Analysis using co-occurring words of 2

Distributions of geotagged tweets of Chiyoda-ku including Ramen obtained in the experiment are shown in Figure 9 for 2017 and Figure 10 for 2018. The interior area of the bold frame in Figure 3 is shown in the table. It is about 80 m measured east–west and about 25 m north–south. This cell is obtained by dividing the maximum value and the minimum value of latitude and longitude into 22 and 7 respectively for geotagging tweets. The closer the color of the cell is to black, the more data are shown.

The data extracted in this experiment were 6,979 in all collection periods. The area of about 540 m east–west and about 540 m north–south in the frame of the heavy line in Figure 3 is a spot called Tokyo Ramen Street, with eight ramen shops. Therefore, as an analytical example using words co-occurring in the target word, Figure 9 and Figure 10 are experimentally obtained results including A="𧰼蔵" co-occurring in the target word and B = "玉".

The ramen shop including A opened in September 2013 and closed in September 2018. The ramen shop including B opened on October 30, 2018.

Therefore, although a difference exists in the number of data, it can be extracted as a spot. However, the closed A information is unsuitable for use as real-time information. The results of analysis considering the time series information are shown in Figure 11. From these results, tweets containing A are not extracted after 2018/9. However, tweets containing B have been extracted since 2018/10. Therefore, by considering the time series in addition to latitude and longitude information, one can omit the old information in addition to extracting new spots. This result confirmed the possibility of realizing real-time information extraction.

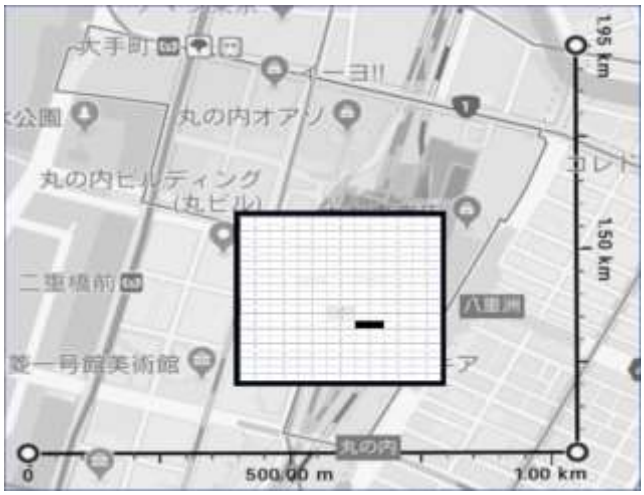


Figure 9: Number of Tweets including “A” Co-occurring in Target Words in Marunouchi 1-chome.

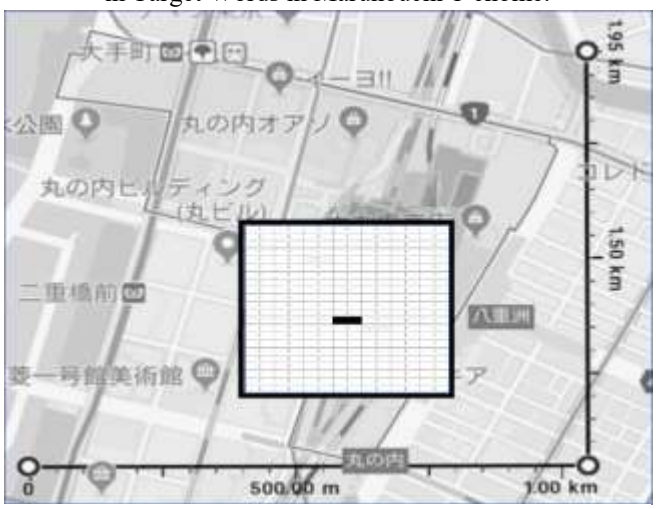


Figure 10: Number of Tweets including “B” Co-occurring in Target Words in Marunouchi 1-chome.

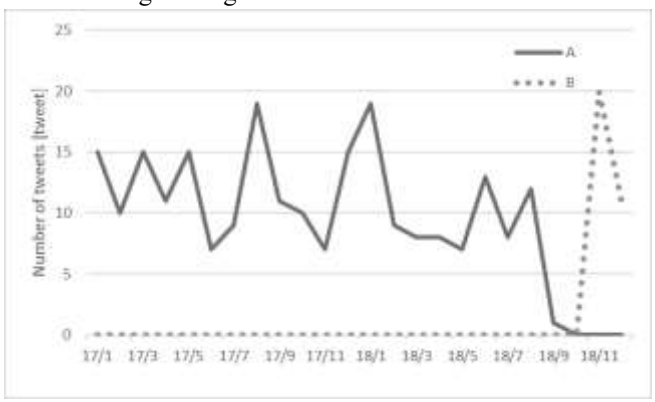


Figure 11: Trends in number of tweets including target words.

4.3.4. Use of the proposed method

The method proposed in this paper assumes the use as a complementary method of real-time analysis using moving average which has been proposed conventionally [2]. As an example, Figure 12 is the estimation result of the cherry blossoms in Tokyo in 2018. Even in Tokyo, where there are a large number of tweets, the number of tweets is small at sightseeing spots alone, such as being limited to Takao. Therefore, we use the proposed method described in Section 4.3.1 to complement the parts that are difficult to estimate in tourist spot units. We will consider combining the recommendation of the spot by the area analysis of the past while making the best estimate in Tokyo unit by this. However, automation of extraction of units to be analyzed and target words are for further study.

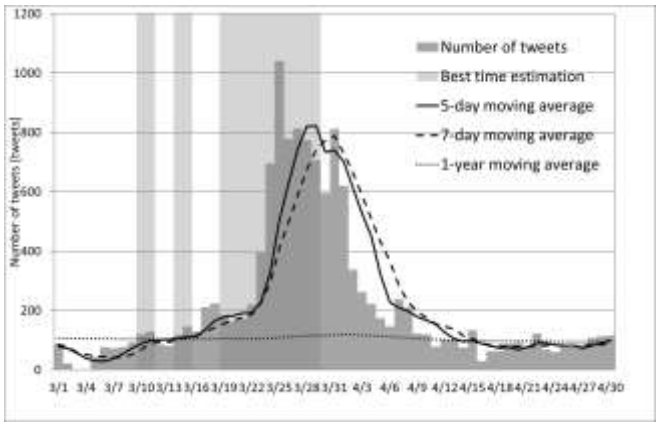


Figure 12: Estimation result of the cherry blossoms in Tokyo in 2018.

5 CONCLUSION

As described in this paper, we evaluated a regional analysis method based on positional information and time series change using tweets with Twitter location information to provide real-time information.

To conduct real-time regional analysis, after proposing a method using geotagged tweets' position information and time series information, we showed experimentally obtained results obtained using that method. Experiment results demonstrated that, even when geotagged tweets were few, spots could be extracted using position information with long-term accumulation. We also confirmed that new spots can be extracted by conducting time series analysis of spot information of position information. Furthermore, using morphological analysis results of tweets, we demonstrated the possibility of analyzing spots, even in densely populated areas with a large amount of information.

Results show that we demonstrated the usefulness of SNS for providing real-time information. In addition, by combining with the conventional method, it is considered that it is possible to perform recommendation on a spot with few geotagged tweets and extraction of a new spot in the target area. Future studies will consider the automation method combined with the conventional method.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Number 18K13254.

REFERENCES

[1] Twitter, It's what's happening, URL < https://Twitter.com/> (2015).

[2] M. Endo, S. Ohno, M. Hirota, D. Kato, and H. Ishikawa, Examination of Best-time Estimation for Each Tourist Spots by Interlinking using Geotagged Tweets, International Journal on Advanced in Systems and Measurements, vol. 10, No 3&4, pp. 163-173, IARIA (2018).

[3] M. Omori, M. Hirota, H. Ishikawa, and S. Yokoyama, Can geo-tags on flickr draw coastlines?, In Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL '14), pp. 425-428, ACM (2014).

[4] T. Sakaki, M. Okazaki, and Y. Matsuo, Earthquake shakes Twitter users: real-time event detection by social sensors, pp. 851–860, WW W 2010 (2010).

[5] A. R. Pratap, J. V. D. Prasad, K. P. Kumar, and S. Babu, An investigation on optimizing traffic flow based on Twitter Data Analysis, 2018 Second International Conference on Inventive Communication and Computational Technologies, pp. 320-325, ICICCT (2018).

[6] Y. Nakaji and K. Yanai, Visualization of Real-World Events with Geotagged Tweet Photos, 2012 IEEE International Conference on Multimedia and Expo Workshops, pp. 272-277, IEEE (2012).

[7] GeoNLP Project, A place name information processing system which maps sentences automatically, URL < https://geonlp.ex.nii.ac.jp/> (2019).

[8] Twitter Developers, Twitter Developer official site, URL <https://dev.twitter.com/> (2015).

[9] Y. Hashimoto and M. Oka, Statistics of Geo-Tagged Tweets in Urban Areas (<Special Issue>Synthesis and Analysis of Massive Data Flow), vol. 27, No. 4, pp. 424–431, JSAI (2012) (in Japanese).

[10] National Agriculture and Food Research Organization, Simple reverse geocoding service, URL <http://www.finds.jp/wsdocs/rgeocode/index.html.ja> (2015).

[11] MeCab, Yet Another Part-of-Speech and Morphological Analyzer, URL < http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html> (2015).

To Detect a Deflation Representing an Event Convergence in Stream Data

Shigeyoshi Ohno*, Takuma Toyoshima* Masaki Endo* and Takuo Kikuchi*

*Division of Core Manufacturing, Polytechnic University, Japan
{ohno, m18303, endou, kikuchi}@uitec.ac.jp

Abstract - In recent years, with the spread of social media such as Twitter, it has become easy to add location information and send then by mobile phones. Since it is possible to observe the real world without using physical sensors, social media have high operational value as social sensors. In this paper, we aim to support decision making for people who are going to visit the specific place where the event or trouble occurred most recently. We propose a method to extract in real time the deflation that represents not only the burst state that shows people's concentration and stagnation but also the continuous flow and dispersion.

Keywords: burst structure, Twitter, real time analysis, human sensor

1 INTRODUCTION

In the modern society in which various things are computerized, there are many means for acquiring various information in real time. One of them, Twitter [1], is a microblogging service called "Tweet" that shares short sentences within 140 characters. It is widely used throughout the world including Japan, and is used by many users as a medium that can post recent information casually. Since it is possible to post location information called geotags easily via a smartphone and post it, it is a social media that can immediately send out where and what is happening. From these characteristics, it is expected as a human sensor for observing the real world without using a high-cost physical sensor [1].

Using human sensors, we want to know the situation in real time even if we are not there. For example, if we can guess the best-time for cherry blossoms and autumn leaves, we don't have to be sorry that cherry blossoms have not yet bloomed or after cherry blossoms have fallen. If public transportation stops and we know that people are crowded, we can avoid that crowding. If we know that the congestion has been eliminated, we can decide that you do not need to avoid it. If we want to go to the Halloween uproar in Shibuya, we want to go to a hurry because it is still uproar. If the Halloween turmoil has settled, it's not interesting to go now. If we do not want to get caught in the uproar, we want to know that the uproar is over. We would like to be able to use human sensors for such behavior judgments.

We want to guess the movement of a person in a specific place in real time so that we can decide whether to go to a certain place or not. It does not make sense to analyze data yesterday's event the next day. It is premised that we are not there. Twitter data should be used to avoid having to deploy expensive sensors. We need a way to know the concentration of people in real time. In this research, we aim to realize this real-time guessing of human motion by analyzing geotagged tweets.

2 RELATED WORK

Kleinberg [2] proposes a method for modeling text stream bursts and extracting structures. This method is based on modeling a atream using an infinite state automaton. The advantage of Kleinberg's approach is that it can represent burst duration, degree, and weight on each topic. Therefore, it is widely used for various applications. However, it is not suitable for real-time burst detection because it is not assumed to analyze immediately for occurrence of a certain event.

Works by Y. Zhu and D. Shasha [3] [4] and the work by X. Zhang and D. Shasha [5] propose a bursting algorithm that efficiently monitors bursts over multiple window sizes. These techniques enable near real-time burst detection by shortening the monitoring interval. However, in order to do so, it is necessary to monitor the number of occurrences of events at regular intervals, and data must be stored even if no events have occurred.

Ebina et al. have proposed a method for detecting bursts in real time [6] [7]. The detection is performed in real time by determining whether or not each event (each tweet posting) is burst. In addition, the amount of calculation is reduced by compressing data held at the time of occurrence of concentrated events, and burst detection with high real-time property is realized. However, it is not clear whether the burst state is continuing or the burst is immediately ended only by the burst occurrence.

Endo et al. use the moving average to make a full-fledged decision [8] [9]. This method is successful not only in the occurrence of burst but also in detection of burst state continuation and convergence. However, since the frequency of the tweet occurrence which fixed window size is used, real-time property will be quantized by window size. To use the frequency of Tweet Posting, we have to set a certain time interval to calculate the frequency. This fixed time interval impairs real-time performance.

In this research, we propose a method to detect convergence of burst state as well as real-time property and burst generation. In reference to the method of Ebina et al., This research guarantees real-time capability by using the tweets posting interval instead of the frequency of tweets posting in a fixed time. In addition, this research aims to detect when the posts decrease and return to the normal state not only in the burst state where tweets post increase rapidly, by combining moving averages with reference to the method of Endo et al.

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

3 OUR PROPOSED METHOD

In order to detect deflation that represents the convergence of an event, it is necessary to capture in real time the phenomenon of people concentrating and spreading at a specific location. We realize real-time performance by monitoring each tweet posting time and its interval. Then, using the Fisher's concept of exact test, the deflation criteria are defined.

3.1 Examination of real-time analysis methods

We experiment using actual tweet information to verify the proposed method. It is not a real-time experiment because it uses collected data. However, since the proposed method does not use information after the determination time, it can perform real-time determination.

The frequency of tweets is not used in our proposal method to take advantage of real-time capabilities. In order to determine the frequency, a fixed time interval must be determined, and this finite time interval impairs real-time performance. Specifically, the tweet time interval is used instead of the frequency with which tweets are posted. Let $c(h, t)$ be the time interval of the tweets posted at time t. The parameter h indicates that it is a time interval from h previously posted tweet. The $c(h, t)$ divided by h is expressed as $ave(c(h, t))$. It is the average value of the tweet occurrence interval from the tweet that occurred h times before time t. If tweet increases, $ave(c(h, t))$ becomes smaller, and if tweet is lost, it takes a large value.

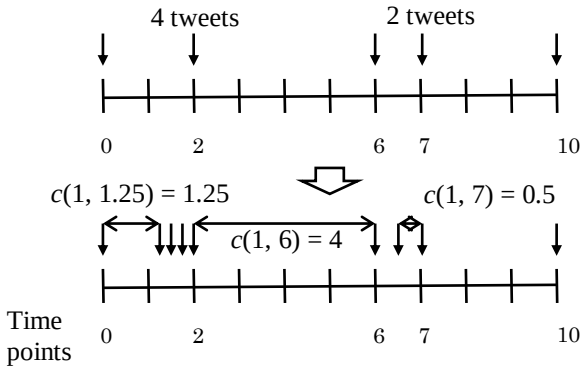


Figure 1: Example of generating function c(h, t).

The timestamp when the tweet is posted is in seconds. If there are many tweets, the timestamps of multiple tweets will be the same. For example, the time points are 2 second and 7 second in Figure 1. Each four tweets and two tweets are posted at the same time. In this case, the value of the function $c(h, t)$ may be zero or not determined. Therefore, as shown in the lower part of Figure 1, the value of the function $c(h, t)$ is made to be uniquely determined by assuming that the tweet is posted by dividing one second. In the case of Figure 1, the values of $c(h, t)$ are as shown in Table 1.

Table 1: Example of the values of function $c(h, t)$.

h	$c(h,1.5)$	$c(h,2)$	$c(h,6)$	$c(h,6.5)$	$c(h,7)$	$c(h,10)$
1	0.25	0.25	4	0.5	0.5	3
2	1.5	0.5	4.25	4.5	1	3.5
3		0.75	4.5	4.74	5	4
4		2	4.75	5	5.35	8
5			6	5.35	5.5	8.25

In the method by Ebina et al. the magnitude of this value is compared to estimate the crowded state whether it is crowded (tweets posting increase) or not. The parameter is used to adjust the comparison. If it does not, it will over-react to a slight change just as a tweet is accidentally posted, and it will be falsely judged that a burst has occurred. By using this parameter to reduce the sensitivity, noise suppression can be achieved.

3.2 Fisher's exact test idea

Fisher's exact test is a statistical test that determines the bias of data divided into two categories.

As an example, consider the case where the efficacy is verified in a clinical trial of a new medicine. The verification is performed on the data obtained by dividing the patient into four according to two exclusive criteria of A and B.

A : Number of patients receiving new medicine.
$\bar{A}$: Number of patients receiving placebo (no-medicinal).
B : Number of patients whose condition recovered after then
$\bar{B}$: Number of patients whose condition did not recover

When data is divided in this way, verification is possible using a contingency table like Table 2.

Table 2: 2 × 2 contingency table.

	B	$\bar{B}$
A	a	b
$\bar{A}$	c	d

By using a contingency table as shown in Table 2, an index s indicating how the disease condition has been improved can be defined as follows.

$$s = ad - bc \quad (1)$$

If the value of s is positive, it means that the new medicine has a “positive” effect on the improvement of the medical condition and has a medicinal effect.

Fisher's exact test is very simple and can be tested if it can be dropped into a contingency table. Using this concept, analysis focusing on individual tweet intervals is performed. Specifically, we propose a method to divide n tweets into 4 regions and analyze the number of tweets in 4 regions. The four areas are generated by setting threshold values on the vertical and horizontal axes as shown in Figure 2. The blue dots in Figure 2 represent individual tweets. The threshold value on the vertical axis is the average value of n tweets, and the horizontal axis divides n tweets arranged in time series into two equal halves.

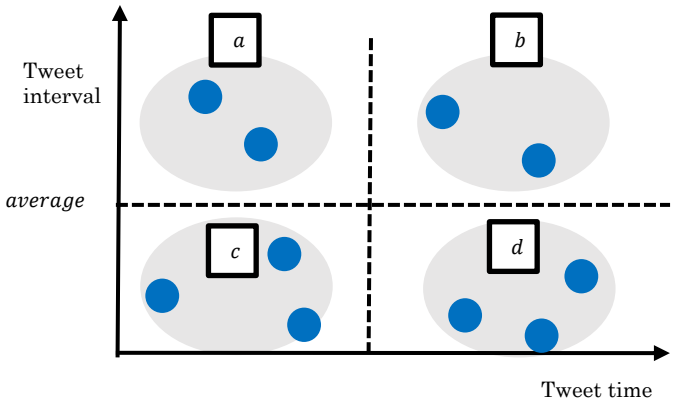


Figure 2: Tweet segmentation diagram using the proposed method.

The a, b, c, and d in the contingency table when analyzing with n pieces are defined as follows.

a: First half of the extracted tweets group and the tweet interval is greater than the threshold (average value)

b: The second half of the extracted tweet group and the tweet interval is greater than or equal to the threshold

c: First half of the extracted tweets group and the tweet interval is smaller than the threshold

d: The second half of the extracted tweet group and the tweet interval is smaller than the threshold

The threshold on the vertical axis representing the tweet interval is the average value of n tweet intervals. When the number of tweets in the region a and the region d increases and s> 0, the tweet interval is narrowed, so that people are concentrated and stagnated, that is, a burst. On the other hand, when the number of tweets in the region b and the region c is increased and s <0, the tweet interval is widened, so that people start to spread and move, that is, deflation. When s = 0, it means that there is no change in the state. In this study, deflation is estimated using s <0 as the deflation criterion.

4 EXPERIMENTS

We experiment using actual tweet information to verify the proposed method. It is not a real-time experiment because it uses collected data. However, since the proposed method does not use information after the determination time, it can perform real-time determination.

Experiments are also being conducted for comparison using the frequency of posting tweets. Those comparison experiments lack real-time capability.

4.1 Data collection

The target event for this experiment was the visit of the General Public to the Imperial Palace after the Accession to the Throne in May 4, 2019. Their majesties the emperor and empress appeared at the balcony of the Chowa-Den Hall to greet the visitors gathered six times. About 140,000 people visited the General Public. Six appearances were held on an hourly schedule starting at 10 am. Participants could enter from the main gate of the Imperial Palace, and the time from 9:30 am to 2:30 pm was the entry time. Analyzes tweets with geotags in the area centered on the Imperial Palace. Those who are in this range are the visit of the General Public who tweet. They will tweet while they are waiting or after they leave. We can imagine that they will not be able to tweet by moving to Chowa-Den immediately before they leave, and that they will refrain from tweeting while they are out. By checking the tweet status, it is possible to estimate the participant's movement (whether waiting or moving).

The tweet to be analyzed was a tweet given the geotag “coordinates”. The geotag “coordinates” is data representing the location where the terminal used for posting a tweet is represented by a single latitude and longitude. Therefore, it can be used as highly positioning data. The extraction period of the tweets was from 00:00:00 on May 4, 2019 to 23:59:59. The extraction range of tweets was empirically determined for the following four latitude and longitude ranges.

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

As a result of filtering under the above conditions, 198 tweets could be extracted. Of these, 116 accounts were tweeted.

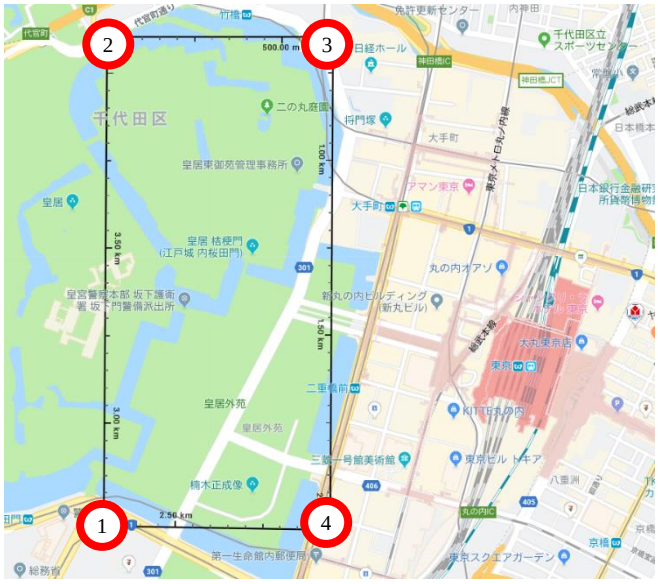


Figure 3: Target Area..

4.2 Results of Comparative Experiments (Method of Endo et al.)

This method uses the frequency of tweets posting. This is in a reciprocal relationship with the Tweet posting interval. Tweet post intervals have high real-time because it is depended on each tweet. On the other hand, in order to calculate the frequency of posting tweets, a certain time interval must be determined. Usually, this time interval is larger than the tweet posting interval, so the real-time performance is low.

The method of Endo et al. uses the moving average of the frequency of posting tweets to estimate the full bloom of

cherry blossoms etc. The method calculates the frequency on a daily basis and examines the difference between the 5-day moving average and the 7-day moving average.

The judgment condition of the best time is when the 5-day moving average becomes larger than the 7-day moving average, and they become larger than the average to the previous year. A comparative experiment was conducted using this condition. However, the frequency of posting tweets is not calculated on a daily basis, but on a 5-minute basis. Unlike the best time to see cherry blossoms and so on, we wanted to know the movement and crowdedness of people, so I made a short time interval.

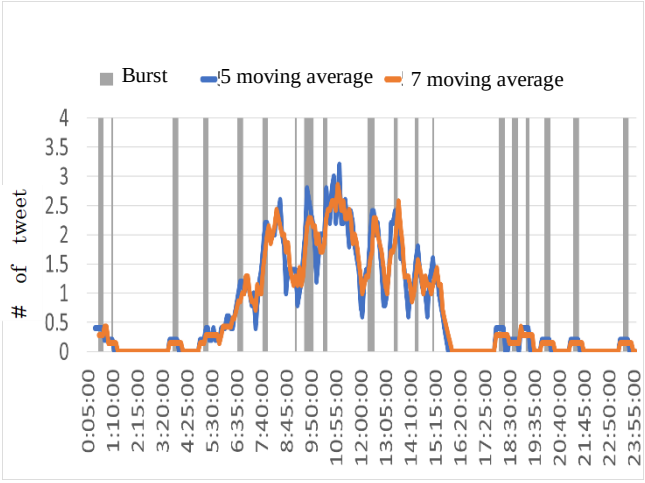


Figure 4: Burst determination using 35-minute moving average (orange line) and 25-minute moving average (blue dotted line).

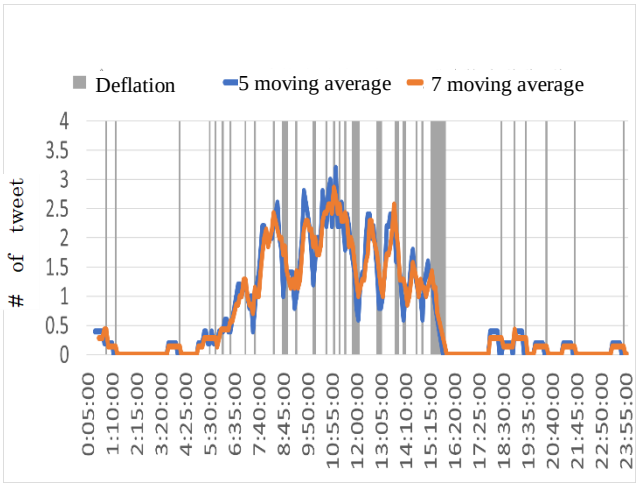


Figure 5: Deflation determination using 35-minute moving average (orange line) and 25-minute moving average (blue dotted line).

The tweet frequency in 5 minutes at time t is expressed by $e(1, t)$. That is, $e(j, t)$ represents the number of tweets posted in $(5 * j)$ minutes from time $(t - 5 * j)$ to time t. The average of e (j, t) divided by j is represented by ave($e(j, t)$) Use the function e to set the burst estimation condition as follows:

$$\text{ave}(e(a, t)) > \text{ave}(e(b, t))$$
$$\text{and} \quad \text{ave}(e(a, t)) > \text{ave}(e(\text{one day}, t)) . \quad (2)$$

As the inverse condition of the burst estimation condition, set the convergence condition of deflation as follows.

$$\text{ave}(e(a, t)) < \text{ave}(e(b, t))$$
$$\text{and} \quad \text{ave}(e(b, t)) < \text{ave}(e(\text{one day}, t)) . \quad (3)$$

Figure 4 shows the judgment under the burst estimation condition, and Figure 5 shows the judgment under the burst convergence condition.

Similar to Endo's 5-day moving average and 7-day moving average, because a = 5 and b = 7, 25-minute moving average and 35-minute moving average will be used. The gray parts in the figure show the time zones that satisfy the judgment conditions.

From Figure 5, the stagnation disappeared and the spread of people was observed at the time after 15:00 when the last appearance. However, both the estimation results by the equations (2) and (3) include many erroneous estimation results. In order to see changes at intervals shorter than 5 minutes, it is expected that a correct judgment can be made by changing to a means having a more real time property.

4.3 Results of Comparative Experiments (Method of Ebina et al.)

The comparative experiment results (Method of Ebina et al.) are shown in Figure 6. The horizontal axis represents time points, and the vertical axis represents tweet occurrence intervals. There is a blue line graph in the figure. This is the average tweet posted interval (average of 5 tweets) representing ave (c (5, t)). Since the left of the figure is the early morning, the tweet occurrence interval becomes large and a line also exists in the upper left. The right side of the figure is midnight, and the tweet interval increases, so there is also a blue line on the upper right.

The crowded (burst) judgment condition of Ebina et al. is the following

$$\text{ave}(c(1, t)) < 0.4 * \text{ave}(c(5, t-1))$$
$$\text{or} \quad \text{ave}(c(2, t)) < 0.4 * \text{ave}(c(5, t-2))$$
$$\text{or} \quad \text{ave}(c(3, t)) < 0.4 * \text{ave}(c(5, t-3))$$
$$\text{or} \quad \text{ave}(c(4, t)) < 0.4 * \text{ave}(c(5, t-4))$$
$$\text{or} \quad \text{ave}(c(5, t)) < 0.4 * \text{ave}(c(5, t-5)) . \quad (4)$$

It can be seen that the deflation judgment is excessive because the interval between the specific tweet and the tweet posted immediately before it is longer than before, and it is determined as a burst. There are times when Appearance time zone and the time when the event ended, and people who were stagnant started moving and were able to judge correctly, but there is a lot of noise to correctly read the spread of people.

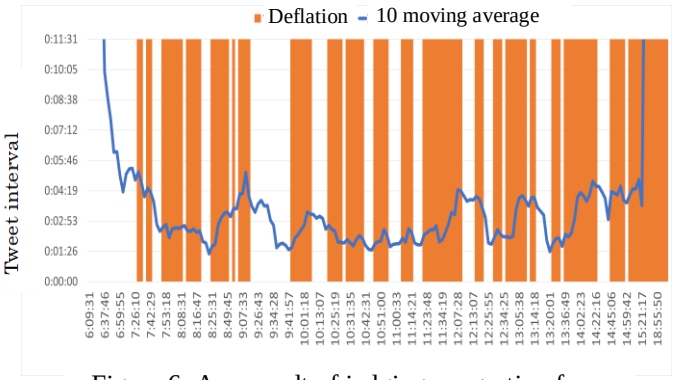


Figure 6: As a result of judging congestion from the tweet information.

4.4 Experimental Results of Our Proposed Method

The results of the proposed method are shown in Figure 7. The broken line in the graph represents the change in the posting time interval of tweets. The gray part shows the deflation judgment result by the proposed method. The time when the value of s in equation (1) is negative was defined as a deflation judgment condition.

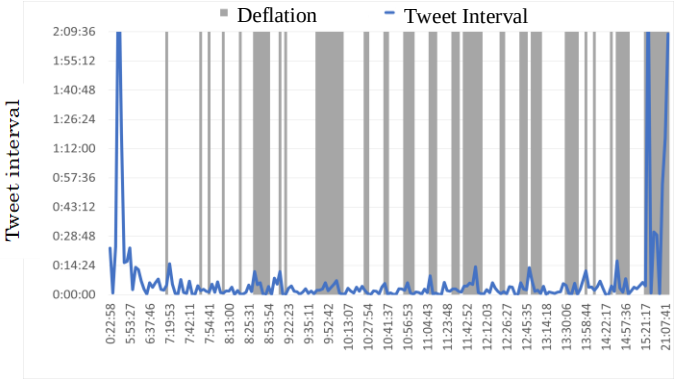


Figure 7: The result of deflation determination by the proposed method.

From the appearance time, we decided the correct time zone that people thought to have moved as shown in Table 3. Some time zones were incorrectly estimated to be deflation, but there were also many time zones that could be estimated fairly accurately, such as around 10:00 and around 12:00. The end of the event after 15:00 can be estimated correctly.

Table 3: The 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	
around appearance	10:40 – 11:10	Move	Deflation
	11:10 – 11:40	Stay	
	:	:	
around appearance	13:40 – 14:10	Move	Deflation
	14:10 – 14:40	Stay	
around appearance	14:40 – 15:10	Move	Deflation
	15:10 – 16:00	Stay	
event end	16:00 –	Converge	Deflation

5 DISCUSSION

In order to quantitatively evaluate the deflation results of this time, a comparison was made using precision, recall, and F value. For correct answer data, the correct answer is determined to be deflation during the time period shown in Table 3. The data to be compared were the experimental results of the technique by Ebina et al. and the experimental results of the technique by Endo et al. Table 4 shows the comparison results.

The reproducibility was not good with the method of Endo et al. using tweet frequency. Although it is a successful method for estimating by day, such as estimate when to see cherry blossoms, it is a disadvantageous method for real-time estimation by minute.

The method of Ebina et al. using the tweet posting time interval has the best recall because of its real-time nature. However, even if the reaction is too good and not deflation, it is erroneously determined to be deflation, and the precision is not necessarily high.

Comparing the relevance rates, the proposed method in this study gave good results. The relevance rate is the best result, and it can be said that our proposed method is a method with good accuracy even when compared with F value.

Table 4: Quantitative evaluation

	Precision	Recall	F-value
Method of Endo et al. Frequency / moving average	48.94%	16.55%	24.73%
Method of Ebina et al. Tweet interval / moving average	42.98%	77.78%	55.37%
Proposed method Tweet interval / number analysis	68.25%	57.33%	62.32%

6 CONCLUSION

We proposed a method to judge the crowded situation in real time by analyzing geotagged tweets. We aimed to detect not only the crowded burst phenomenon but also its convergence, real-time performance is realized. We succeeded in detecting the phenomenon of deflation correctly. And the precision and F-value of the proposed method are the best.

However it has not been completely successful yet. Increase versatility as a deflation estimation method. The method must be adaptable to various events. We must improve the accuracy further.

7 ACKNOWLEDGMENTS

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

REFERENCES

[1] T. Sakaki and Y. Matsuo, Twitter as a Social Sensor : Can Social Sensors Exceed Physical Sensors? Journal of Japanese Society for Artificial Intelligence 27(1), pp.67-74 (2012).

[2] J. Kleinberg, Bursty and hierarchical structure in streams, In Proc. 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.91-101, ACM(2002).

[3] Y. Zhu and D.Shasha, Efficient Elastic Brust Detec-tion in Data Streams, Proc. 9th ACM SIGMOD Inter-national

Conference on Knowledge Discovery and Data Mining, pp.336-345, ACM(2003).
[4] D. Shasha and Y. Zhu, High Performance Discovery in Time Series, Techniques and Case Studies (Monographs in Computer Science), Springer-Verlag(2004).
[5] X. Zhang and D. Shasha, Better Burst Detection, Proc. 22nd International Conference on Data Engineering, pp.146-149, IEEE computer Society(2006).
[6] R. Ebina, K. Nakamura and S. Oyanagi, A Proposal for a Real-Time Burst Detection Method, DBSJ Journal, Vol.9, No.2, pp.1-6(2010).
[7] R. Ebina, K. Nakamura and S. Oyanagi, A Proposal for a Real-time Burst Analysis Method, IPSJ TOD, Vol.5(3), pp.86-96(2012).
[8] M. Endo, Y. Shoji, M. Hirota, S.Ohno and H. Ishikawa, On Best Time Estimation Method for Phenological Observations Using Geotagged Tweets, IWIN2016, pp.205-210 (2016).
[9] M. Endo, M. Hirota, S.Ohno and H. Ishikawa, Best-Time Estimation Method by Region and Tourist Spot using Information Interpolation, IWIN2016, pp.209-216 (2017).

Session 5:
Data Models
(Chair: Takuya Yoshihiro)

Training Data Generation Method for Deep Learning Using Cycle-GAN

Tsukasa Kudo[†]

[†]Faculty of Informatics, Shizuoka Institute of Science and Technology, Japan
kudo.tsukasa@sist.ac.jp

Abstract - Image recognition using deep learning has been applied in various fields, and high recognition accuracy has been achieved. However, there is a problem that a large amount of training data has to be prepared for deep learning, and it often becomes obstacles to applying deep learning. For this problem, I had conceived to use computer graphics (CG) images instead of actual pictures and developed the test data generator by using CG, then I performed some experiments to evaluate its effectiveness. As a result, I found there was a problem to achieve the accuracy same as actual pictures were difficult because it was hard to adjust the color tone of CG images. On the other hand, recently, Cycle-consistent Adversarial Networks (Cycle-GAN) has been proposed, and it has been shown that an image could be translated into a fake image having different color and texture automatically. This suggests that actual pictures can be translated into fake CG images, that is, the above color tone problem can be solved. In this paper, even for the models of deep learning trained with CG images, I show the accuracy of image recognition can be improved by using actual pictures translated by Cycle-GAN model.

Keywords: Deep learning, Cycle-GAN, Regression, Stock-taking, Computer graphics, Image recgition

1 INTRODUCTION

In recent years, the effectiveness of deep learning has been recognized, and its various models have been proposed and studied [3], [2]. In particular, the accuracy of image recognition has been rapidly improved since its effectiveness has been shown by the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [7], [13]. As a result, deep learning has become to be applied to image processing in various fields such as face recognition, medical image analysis, epidemic prevention of plants, and so on [14], [5], [11]. One of the reasons for this can be pointed out that various images have been shared in cloud servers as big data in recent years, and it has become easy to obtain a large number of images as training data. On the other hand, in the field that targets peculiar objects, preparing a large amount of training data often takes a lot of works and becomes an obstacle to applying deep learning.

As such the case, I have been dealing with the issue of stock-taking in a machine factory. Though most parts of machine factories are usually stored in bulk containers, the number of parts in bulk containers cannot be counted externally by human vision. So, for stock-taking, since it is necessary to take out and count parts, there is a problem that a large man-hour is required. For this problem, I conceived a method to use deep learning, and in this method, the stock-taking was performed inventory estimation by using image recognition. As a result, I confirmed that practical accuracy could be achieved by using a regression model of deep learning [9]. On the other hand, deep learning requires a large amount of training data. So, in this study, we prepared 1,600 original images, then we increased them to 8,000 images as training data by padding. In the case of parts in the actual factory, this becomes a task of manually repositioning the parts and taking pictures. However, many parts are heavy and there are often thousands of bulk containers in a factory. Therefore, it is not a practical method to do such work for every bulk container.

Here, I had already developed a training data generator using computer graphics (CG) to generate a large amount of training data automatically and confirmed its effectiveness for the multi-class classification model of deep learning [8]. So, we trained the regression model with CG images generated by this generator and performed the comparative evaluations of estimation accuracy between using pictures (actual images) and CG images. However, in the case of using the actual images, the accuracy deteriorated significantly. Here, parts of the factory are processed based on the drawing of machining. Therefore, even for CG, by using the same drawing, the high accuracy in the shape of the part model could be achieved easily.

However, since the color tone of parts is affected by the texture and illuminations, it was necessary to adjust it with trials and errors. And, as the target accuracy of this color tone was increased, the man-hours for the adjustment increased rapidly. As a result, it was difficult to create CG images that perfectly match the actual image in color tone. And, I also found that this difference in color tone was the cause of the deterioration in inventory estimation accuracy [9]. Furthermore, in the actual factory, since there is a change in the environment such as brightness difference between day and night, it is expected that there is further deterioration.

On the other hand, recently, the generative adversarial network (GAN) has been proposed, and it has become possible to generate automatically a large number of fake images almost like genuine [4]. And, various GAN models have been proposed [10]. For example, by using Cycle-consistent Adversarial Networks (Cycle-GAN) model, which is trained by different image groups such as zebras and horses, images of one group (zebras) are automatically translated into fake images of another group (horses). And, as for some kinds of objects, it has been shown that images can be translated into high accurate fake images using this model [16]. This suggests that even with the above-mentioned parts of the factory, it may be possible to generate fake CG images, which are indistinguishable with original CG images, from actual images.

Figure 1: Inventory shelves of bulk container

That is, the above-mentioned problem on color tone may be solved.

The aim of this study is to verify the feasibility of improving the estimation accuracy of the above-mentioned regression model in the inventory estimation by using Cycle-GAN. That is, the regression model is trained by CG images generated by the training data generator. On the other hand, the actual images of the bulk containers are translated to fake CG images by using Cycle-GAN, and inventory estimations are performed using these fake images. By this translation, it is expected to be effective not only for eliminating color tone differences between the actual and CG images but also for adapting to changes in the environment in the actual factory.

To evaluate this effect, in this paper, I perform comparative evaluations of inventory estimation accuracy for the bulk container using CG images, fake CG images and actual images by the regression model trained with CG images. And, even in the case where training data is prepared by the CG based training data generator, I show that the estimation accuracy can be improved by translating actual images to fake CG images.

The remainder of this paper is organized as follows. Section 2 shows related works and the aim of this study and Section 3 shows the method for estimating inventory with the training data generator and Cycle-GAN model. Section 4 shows the evaluations of effect in the case of using fake CG images translated by Cycle-GAN. In Section 5, I discuss the evaluation results, and I conclude in Section 6.

2 RELATED WORKS AND AIM OF STUDY

The parts inventory in the machine factory is managed based on the theoretical inventory planned by using the production management system. However, there are often gaps between the actual and theoretical inventories due to defects, work errors, work delays and so on. So, it is necessary to carry out stock-taking to grasp the actual inventory for maintaining appropriate inventory. However, since most parts of the factory are stored in bulk containers shown in Fig. 1, they cannot be counted from the outside visually. So, it is necessary to take out and count the parts, which increases the load of inventory work.

On the other hand, in recent years, with the progress of deep learning, the accuracy of image recognition has been improved, and it has been applied to various fields [7], [13], [14], [5], [11]. So, I conceived the method to estimate inventories by using pictures taken from the outside of the bulk containers, and I have evaluated its effect. Firstly, I evaluated the case of using the multi-class classification model of deep learning and showed that it could achieve a certain estimation accuracy. In addition, I showed the way to apply this method to actual inventory management [6]. As a result, though there were estimation errors in this method, I showed that out of stock could be restrained by increasing safety stock according to the range of error. In addition, by comparing with theoretical inventory, the accuracy could be more improved. Incidentally, in the multi-class classification model, inventory quantities are treated as discrete classes, and an estimated inventory of bulk container picture is classified into one class (quantity) [2].

However, I found that preparing training data was an enormous workload through that study. So, I developed a training data generator using CG and showed that a certain accuracy could be achieved with the training data generated by this [8]. And, I also found the followings. Since parts were processed based on the drawing of machining, the shape of the CG part model also could be created relatively easily from the drawing. Conversely, as for the color tone of the model, trials and errors were necessary to adjust it according to the texture of part material and illumination environment. Incidentally, since parts were processed into various shapes from limited types of materials such as iron and aluminum, the number of types of textures was very small compared to the types of parts.

On the other hand, in the multi-class classification model, it is necessary to prepare training data for each inventory quantity. So, in the case where inventory quantity fluctuates over a wide range, it is necessary to prepare a large amount of training data. Therefore, I evaluated the accuracy using a regression model [9]. Regression model estimates inventory quantity as continuous quantity [2]. As a result, in the case of using actual images, high estimation accuracy was obtained compared to the case of the multi-class classification model. In addition, I confirmed that similar accuracy could be obtained for plural parts by estimating with CG images using models trained with CG images. However, in the case of estimating with actual images using these models trained with CG images, I found that the estimation accuracy is significantly deteriorated due to the difference in color tone between the actual and CG images. Furthermore, in the actual factory, there is also a problem that the color tone of the actual image slightly changes depending on such as day and night and the weather.

By the way, in 2015, by using DeepDream, which is an image processing method using deep learning, it was shown that images can be generated automatically. Since then, studies on automatic image generation and image style conversion by applying deep learning have been actively conducted [12], [2]. Among them, GAN is a method to generate a fake image just like a genuine image and is composed of a generator network (generator) that generates a fake image and a discriminator network (discriminator) that distinguishes the fake image genuine or fake [4]. By making them train mutually, it

Figure 2: Image translation from zebra to horse [16]

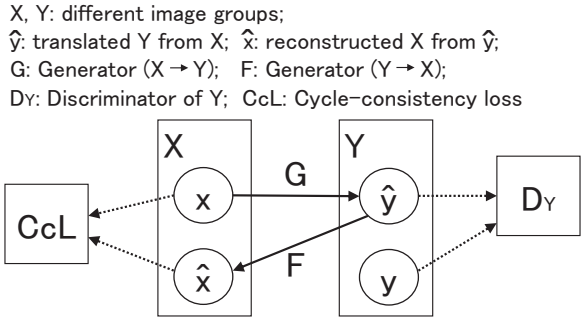


Figure 3: Structure of Cycle-GAN model

is possible to generate a fake image close to the genuine image. And, various methods for generating fake images have been proposed by applying this GAN [10].

Among them, Cycle-GAN translates an image of one group into an image of another group. For example, as shown in Fig. 2, an image of zebras is translated to a fake image of horses by training a Cycle-GAN model with the image group of zebras and horses [16]. Figure 3 shows a structure of Cycle-GAN model, and X and Y shows different image groups corresponding to such as zebra and horse image group respectively.

In Fig. 3, image x of X are translated to fake image $\hat{y}$ of Y by generator G; discriminator D_Y determines whether $\hat{y}$ is a genuine or fake image of Y. Furthermore, $\hat{x}$ is reconstructed from $\hat{y}$ by generator F, and Cycle-consistency loss, which is a loss between x and $\hat{x}$, is evaluated. Here, Fig. 3 shows the case of translating X to a fake image of Y. Similarly, images of Y are also translated to fake images of X, and discrimination and cycle-consistency loss evaluation are performed. In this way, similar to GAN, by making generator and discriminator train mutually, images of one group become to be translated into high-precision fake images of another group.

Furthermore, since the Cycle-GAN model has the structure shown in Fig. 3, it can perform automatically such image translation between unrelated image groups X and Y. Therefore, by applying the Cycle-GAN model to the above-mentioned inventory estimation, it is expected to be automatically translated actual images of bulk containers taken in various environments into fake CG images generated by the training data generator. However, though some application cases of Cycle-GAN have been shown [1], [15], they were about the image to image translations; and I could not find the application case to image recognition.

This study aims to solve the problem in the color tone difference between actual images and the CG images which is

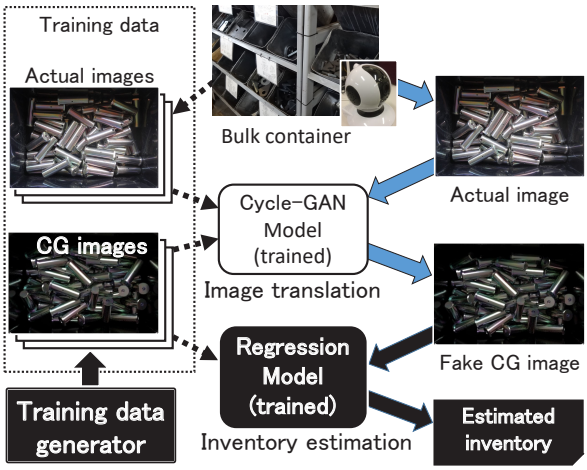


Figure 4: Dataflow of inventory estimation system

generated by the training data generator, and to verify the feasibility of an inventory estimation system using CG images as training data. So, I show that estimation accuracy can be improved by translating the actual images into fake CG images by using Cycle-GAN, even in the case where quantity is estimated by using the regression model which is trained with the CG images.

3 INVENTORY ESTIMATION PROCESS USING TRAINING DATA GENERATOR

Figure 4 shows the data flow of the inventory estimation system mentioned in Section 2, in which Cycle-GAN is used to improve estimation accuracy. As shown in the dashed box in Fig. 4, training data for Cycle-GAN model is prepared as followings. Actual images are created from pictures of the factory bulk containers; CG images are generated by the training data generator described in Section 2. And, this system uses the regression model trained by CG images, as shown in the black hatched round box. This regression model consists of convolutional layers with pooling layers and fully connected layers as shown in Fig. 5. In addition, the mean square error (MSE) is used for the loss function.

Also, as shown in Fig. 4, in this system, the actual image of each bulk container is extracted from the picture of the camera installed around the containers, and the inventory is estimated from this image. For this actual image, to suppress the increase in the inventory estimation error due to the color tone difference between the actual and CG image mentioned in Section 2, it is translated to a fake CG image by using the Cycle-GAN model. And, the above-mentioned regression model estimates inventory by using this fake CG image.

As for training data of the regression model, the shape of the part is also important to maintain estimation accuracy. Also, since one camera monitors many bulk containers, it is necessary to create training data images not from just above direction but from the direction of the camera for each bulk container. In these respects, with the training data generator, the shape of the CG part model can be easily created based on the drawing of machining of the target part, and the camera position of CG can be designated at rendering based on the ac-

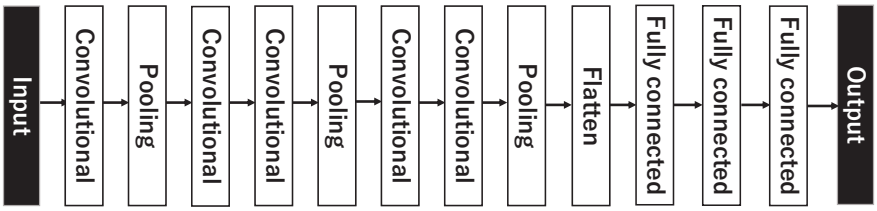
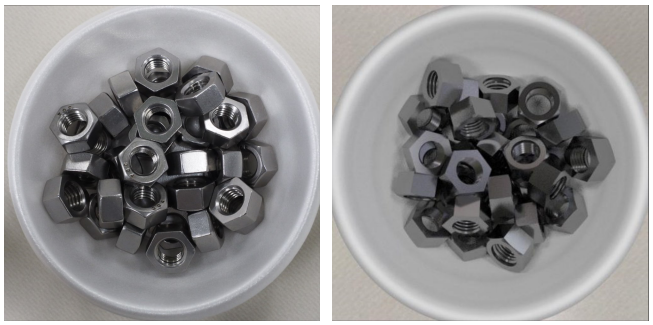


Figure 5: Structure of regression model



(1) Actual image (2) CG image

Figure 6: Target nut images

tual camera position. That is, the training data generator can easily generate a large amount of training data automatically, except for the problem of the color tone of the part model.

Therefore, the Cycle-GAN model is used to translate the color tone of part, which mainly depends on part material. Here, as mentioned in Section 2, since there are relatively few types of part materials, training of the Cycle-GAN model is conducted just for each of these materials. That is, for the training data, CG images are generated by the training data generator; actual images are created by collecting pictures of parts made of the same material. In addition, as mentioned in Section 2, for Cycle-GAN training data, it is not necessary to associate between these actual and CG images. So, it is possible to accumulate the actual images without investigating the number of parts.

4 EXPERIMENTS AND EVALUATIONS

4.1 Experimental Environment

The experimental environment to evaluate the effect of translating actual images to fake CG images by using Cycle-GAN in inventory estimation is similar to the inventory estimation system shown in Fig. 4. Firstly, for target parts of evaluations, I used images of nuts placed on the bowl, which were taken from just above as shown in Fig. 6. For example, Fig. 6 (1) shows an actual image created from the picture of the nuts; (2) shows a CG image generated by the training data generator. For each of them, I prepared 100 images for each number of nuts from 5 to 80 for every 5, namely, 16 classes, total 1,600 original images. Then, I divided them into 1,200 training data, 400 test data.

Next, I show the structure of the experimental system in Fig. 7. The Cycle-GAN model was trained by using both actual and CG images. This training was performed as batch

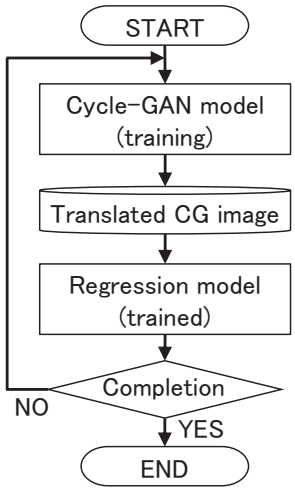


Figure 7: Structure of experimental system

processing, and the number of batch processing per epoch was 1000 times. Here, every 200 times of batch processing, the actual images prepared separately were translated into fake CG images, and inventory estimation by the regression model was performed. Classes of these actual images were similar to the training data, that is, it composes 16 classes for the number of nuts from 5 to 80 for every 5. And I prepared 50 images for every class, namely 800 images of the total. In addition, in order to evaluate the estimation error automatically, the number of nuts in each image was added as a correct label.

In addition, this experimental system was implemented on a PC with Windows 10 using Python and Keras, and Tensorflow was used for the backend of Keras; OpenCV was used for image conversion, and the images were converted into the size of 128×128 and used for training and estimation. Incidentally, for the training data generator, I modeled a nut and bowl using Blender which was a 3DCG modeling tool and placed the nuts on the bowl by using its physical simulation function. And, I automated Blender's physical simulation and rendering using Python.

Also, for the training of the Cycle-GAN model, I set $\lambda = 10.0$, which was the strength of the cycle-consistency loss against the discriminator D_Y in Fig. 3 and D_X which was the discriminator for generator F; As for Adam optimizer, its argument was set as $Adam(0.0003, 0.5)$. Similarly, for the regression model, the reduction rate of learning rate was 0.1, the minimum learning rate was 10^{-10}, Dropout was not used, and the output dimensions of fully connected layers were 128 except the last layer, and the best model in the training transition was saved. Note that the output dimension of the last

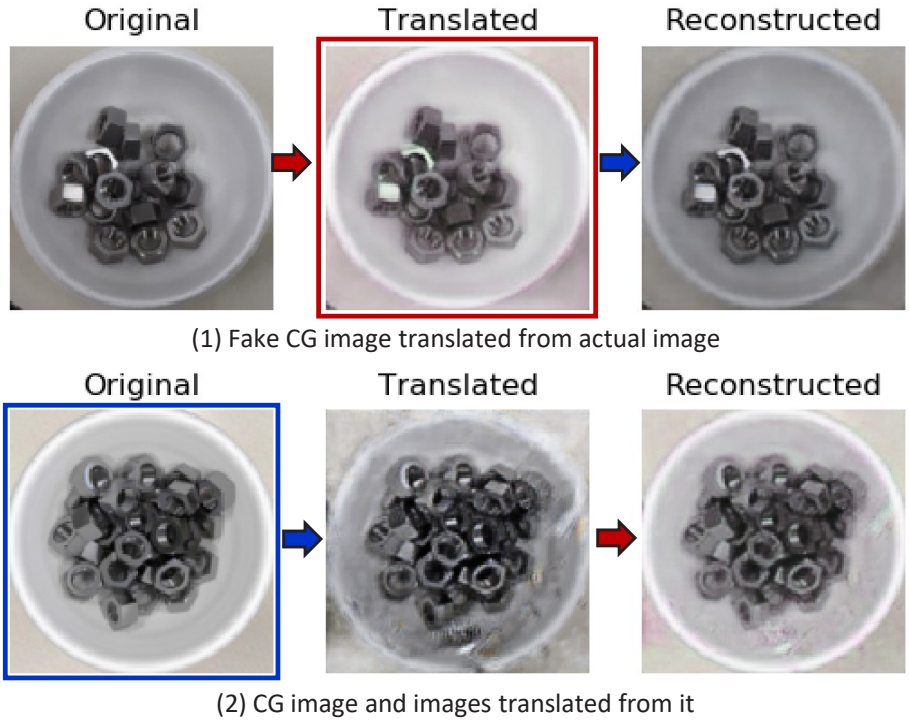


Figure 8: Generated images by cycle-GAN model

layer of a regression model is one.

4.2 Evaluations

Figure 8 shows an example of input and output images of the Cycle-GAN model when the MSE, namely the loss of the regression model to estimate, became the smallest. Figure 8 (1) shows, from the left, the actual image, the fake CG image translated from this actual image, and the actual image reconstructed from this fake CG image. Each corresponds to x, $\hat{y}$, $\hat{x}$ in Fig. 3 respectively. Similarly, (2) shows a CG image, its fake actual image, and reconstructed CG image. That is, inventory estimation was performed by the regression model using the fake CG image at the center in (1). The original image at the left in (2) is a CG image, and the translated fake CG image in (1) was generated as a fake image of this CG image.

Next, to evaluate the effect of images translated by Cycle-GAN on inventory estimation, namely the effect of fake CG images, I performed a comparative evaluation using CG images, fake CG images, and actual images. Figure 9 shows the evaluation results for these kinds of images by the MSE. The MSE of using fake CG images was improved about 2.8 times compared to the case of using actual images. However, it was deteriorated about 9.2 times compared to the case of using CG images.

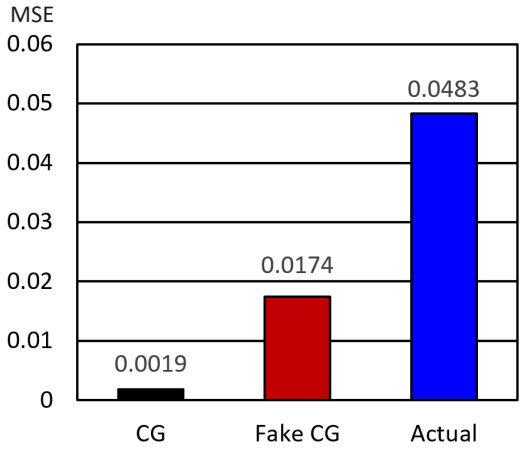


Figure 9: Comparative evaluation of MSE of estimated inventory

5 DISCUSSIONS

In the previous study, I evaluated the estimation accuracy of parts inventories in bulk containers by the regression model using the following data. CG data generated by the data generator was used for training the model; Actual images were used for estimating inventory by the model. As a result, although the efficiency of training data generation improved, the problem occurred that the estimation accuracy deteriorated. In this paper, to solve this problem, I am trying to investigate the feasibility of the method of translating actual images to fake CG images using the Cycle-GAN.

Firstly, I performed the comparative evaluations of the estimation accuracy with and without color translation by using the Cycle-GAN model. As a result, as shown in Fig. 9, it was confirmed that this color translation was effective for improving the estimation accuracy. By the way, as shown in Section 2, I have previously shown the way to apply to an actual factory by increasing the safety stock and collating with the theoretical inventory. Furthermore, since I have also previously

confirmed that the types of materials of parts were limited as mentioned in Section 2, it is expected that one model of Cycle-GAN can be applied to plural parts. Therefore, considering the trade-off between the increased cost of safety stock and the efficiency of training data generation, I consider that there is a field that this method can be applied even with the current accuracy.

However, the MSE of inventory estimation using fake CG images was about 9.2 times that of using CG images as shown in Fig. 9. To solve these problems, I consider the model is necessary to make the loss function of the Cycle-GAN model reflect the above-mentioned MSE, which is the loss function of the regression model. By using this model, the Cycle-GAN model which is optimal for the regression model to estimate inventory can be trained. As a result, it is expected to close the accuracy with Fake CG to the case of CG shown in Fig. 9. The development of this model is my further study.

6 CONCLUSION

For inventory estimation of bulk containers using the regression model of deep learning, training data can be efficiently prepared by training data generator which uses CG. However, there was the problem that the estimation accuracy deteriorates due to the difference in the color tone between CG and actual images.

For this problem, in this study, I evaluated the method of generating fake CG images by translating the color tone from actual image by using Cycle-GAN. Concretely, I performed the comparative evaluations of the estimation accuracy using CG images, fake CG images, and actual images. As a result, it was confirmed that the accuracy of inventory estimation could be improved by using these fake images rather than using the original actual images. On the other hand, the estimation accuracy of using fake CG images was lower than that of using CG images.

To improve the accuracy of using fake CG images, I consider that it is necessary to reflect the loss of the regression model, namely MSE, into the loss function of Cycle-GAN model. This is a future study.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 19K11985, and a research grant of Toukai Foundation for Technology.

REFERENCES

[1] A. Bansal, S. Ma, D. Ramanan, and Y. Sheikh, “Recycle-gan: Unsupervised video retargeting,” Proc. European Conference on Computer Vision (ECCV), pp. 119–135 (2018).

[2] F. Chollet, “Deep learning with python,” Manning Publications Co. (2017).

[3] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, “Deep learning,” MIT press (2016).

[4] I. Goodfellow, “Nips 2016 tutorial: Generative adversarial networks,” arXiv preprint arXiv:1701.00160 (2016).

[5] H. Greenspan, B. van Ginneken, and R. M. Summers, “Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique,” IEEE Transactions on Medical Imaging, Vol. 35, No. 5, pp. 1153–1159 (2016).

[6] T. Kawanaka, and T. Kudo, “Inventory Satisfaction Discrimination Method Utilizing Images and Deep Learning,” Procedia Computer Science, Vol. 126, 937-946 (2018).

[7] A. Krizhevsky, I. Sutskever, G.E. Hinton, “Imagenet classification with deep convolutional neural networks,” Advances in neural information processing systems, pp. 1097–1105 (2012).

[8] T. Kudo, and R. Takimoto, “A proposal of deep learning training data generator for image recognition,” Proc. Int. Wokshop on Informatics (IWIN2018), pp. 119–126 (2018), http://www.sist.ac.jp/˜kudo/research/IWIN2018_119_126.pdf (reffered June 5, 2019).

[9] T. Kudo, “CG Utilization for Creation of Regression Model Training Data in Deep Learning,” Procedia Computer Science, (in press).

[10] E. Linder-Norén, “Keras implementations of Generative Adversarial Networks”, https://github.com/eriklindernoren/Keras-GAN (reffered June 5, 2019).

[11] S. P. Mohanty, D. P. Hughes, and M. Salathé, “Using deep learning for image-based plant disease detection,” Frontiers in plant science, Vol. 7, Art. 1419 (2016).

[12] A. Mordvintsev, C. Olah, M. Tyka, “DeepDream - a code example for visualizing Neural Networks,” Google AI Blog, https://mng.bz/xXlM (reffered June 5, 2019).

[13] M. Verhelst, and B. Moons, “Embedded Deep Neural Network Processing: Algorithmic and Processor Techniques Bring Deep Learning to IoT and Edge Devices,” IEEE Solid-State Circuits Magazine, Vol. 9, No. 4, pp. 55–65 (2017).

[14] Y. Wen, et al., “A discriminative feature learning approach for deep face recognition,” European conference on computer vision, Springer, pp. 499–515 (2016).

[15] L. Wei, S. Zhang, W. Gao, and Q. Tian, “Person transfer gan to bridge domain gap for person re-identification,” Proc IEEE Conf. on Computer Vision and Pattern Recognition, pp. 79–88 (2018).

[16] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” Proc. IEEE Int. Conf. on computer vision, pp. 2223–2232 (2017).

Proposal and evaluation for property verification for Java functions with recursive data structures by SAW

Rin Karashima†, Satoshi Harauchi‡, Shinpei Ogata*, and Kozo Okano*

†Shinshu University, Japan
‡Mitsubishi Electric Corporation Advanced Technology R&D Center
* Shinshu university, Japan
†19w2036c@shinshu-u.ac.jp
‡Harauchi.Satoshi@bc.MitsubishiElectric.co.jp
* {okano, ogata}@cs.shinshu-u.ac.jp

Abstract - In software development, there is often an opportunity to recreate a program that behaves as same as the program you created in another programming language. At this time, in order to avoid unexpected bugs, it is necessary to verify the equivalence of the behavior of the new and old programs and whether the program behaves as intended. One of the common verification methods is human verification using test cases, but there are problems in terms of verification omission and cost. SAW (*software Analysis Workbench*) has been released as a verification tool for programs based on formal methods. In SAW, it is possible to verify whether the program to be verified satisfies the verification property by described in a verification script called SAW Script. However, verification using SAW is not a *Silver Bullet* because it's under development technology and methods of verification are not organized. We do not know for example, ability to equivalence check between functions written in Java with recursive data structures. This paper investigates the possibility of equivalence check between functions written in Java with recursive data structures. We show the positive results for proposing a concrete method for equivalence check between functions written in Java with recursive data structures.

Keywords: SAW, equivalence checking, refactoring Java, recursive data structures

1 INTRODUCTION

Both of Java and C Programs are widely used for modern information systems. Sometimes, change of the environment lets engineers develop new software codes with the same functions to the old codes work on the old environment. For example, c programs working on an old environment are changed to a new Java programs working on a new environment, where these programs behave the same functions.

In such cases, programmers have to ensure that the revised Java program also preserves the behavior of the old C version of the program.

Yet another example, programmer might have to develop new C programs for some embedded system where CPU and memory resources are limited, based on an existing Java programs working on a new rich environment. Also, in such a case, programmers have to ensure that the revised C program also preserves the behavior of the old Java version of the program.

Usually, regression testing is performed to check whether the both of the versions have the same behavior. In the above situation, however, usually we cannot use the same regression test-suits due to different programming languages.

Formal Approach Techniques might help for such a situation. These techniques will find potential bugs or firm confidence on the conformance with adequate efficiency. We call this kind of verification formal conformance verification (FCV).

Recent tools, however, do not fully support program dealing dynamic data structures especially recursive data structures.

For example, in our previous research [15] finds the possibility of FCV for C programs with recursive data structures. We resolve this problem by using idea of bound model checking.

However, for Java programs, we don't know the possibility. The reason is summarized as follows.

- SAW retrieves information used for verification from byte codes of C or Java. These byte codes are LLVM and jvm. The difference of the codes affects the checking algorithm used in SAW.

- Java is based on Java virtual machine while C has no reference machine to interpret.

In order to avoid the halting problem, in other words, if verification is performed without defining the end of the linear list structure, problems such as failure to complete the verification without completing the recursion and state explosion will occur, the method is based on bounded model verification technique [1], [2]. We also perform experimental evaluation using SAW (Software Analysis Workbench)[3]. SAW is a recent formal verification tool.

From the results of the experiments, we find that it is partially possible to perform verification of function dealt with recursive data structures in Java. This observation supports positive possibility of FCV among C and Java with recursive data structures.

The rest of this paper organized as follows. Section 2 gives preliminaries. Section 3 describes the proposed method. Section 4 and 5 give experimental evaluation. Section 6 discusses the results. Finally, Section 7 summarizes this paper.

2 PRELIMINARIES

2.1 Definition of Equivalence between Functions

For given two program functions f and g with the same parameters, we call that functions g and f are equivalent if expression (1) holds.

$$\forall x : f(x) = g(x) \tag{1}$$

The parameter x can be easily extended to a parameter vectors with the same signatures. Here, a signature is a list of types corresponding to each element of parameters.

Please note that we don't care the program languages used for implementing functions f and g.

This paper usually deals with the case that f is implemented in Java, while g is implemented in C.

2.2 SAW

SAW (*software Analysis Workbench*) [3] is an open source tool developed by Galois Inc. that allows formal verification of program functions written in C, C ++, Objective-C, and Java. Figure 1 illustrates the architecture of SAW. SAW reads a binary file llvm and jvm for a program written in C or Java. It uses SAT/SMT solvers to check the satisfiability of the verification properties specified in the verification script. The verification script is called SAW Script.

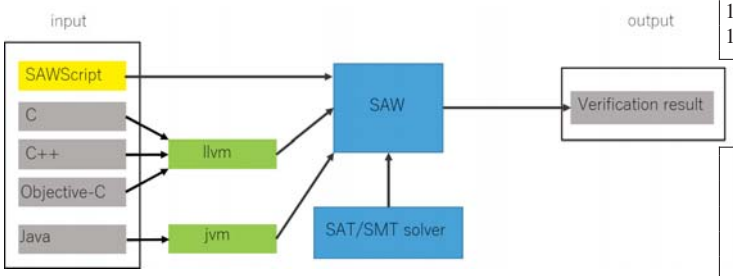


Figure 1: SAW Architecture

Here, we explain an example of verification using SAW for functions written in C and Java. Listings 1 and 2 show functions that return the doubled argument written in C and Java, respectively. Listing 3 is a SAW Script that proves the equivalence of these functions using `java_extract`. The description at Line 10 defines the equivalence on both of return values of the two functions as the verification property. The proof is performed using a verification engine called ABC [13] by the description of Line 11. This script can be run on a terminal to prove it. If verification is successful, only a success message is displayed. If it fails, it shows a concrete example of an argument that fails verification.

SAW Script shown in Listing 4 uses `java_verify` to prove that the function shown in Listing 2 satisfies the verification property described in `function_spec`. Line 2 shows that the argument x of the function to be verified is given as integer. Line 3 indicates that the return value of the function to be verified is always $x * 2$ in 32-bit integer. Line 4 lets SAW use ABC, which is the default inference engine of ABC. SAW specifies arguments and return value of the function under verification. Users describe the above and the verification property in SAW Script.

Listing 1: Mul.c

```
1  #include <stdint.h>
2  int old_function(int x){
3    return x * 2;
4  }
```

Listing 2: Mul.java

```
1  class Mul{
2    int new_function(int x){
3      return x << 1;
4    }
5  }
```

Listing 3: SAWScript

```
1  // Load C function
2  c_mul <- llvm_load_module "Mul.bc";
3  c_func <- llvm_extract c_mul "old_function"
       llvm_pure;
4
5  // Load Java function
6  java_mul <- java_load_class "Mul";
7  java_func <- java_extract java_mul "
       new_function" java_pure;
8
9  // Prove
10 let thm = {{ \x -> c_func x == java_func x}};
11 prove_print abc thm;
```

Listing 4: SAWScript

```
1  let function_spec : JavaSetup () = do {
2    x <- java_var "x" java_int;
3    java_return {{ x * 2 : [32] }};
4    java_verify_tactic abc;
5  };
6
7  java_mul <- java_load_class "Mul";
8  java_verify java_mul "new_function" []
       function_spec;
```

2.3 SAT/SMT Solvers

SAT (SATisfiability problem) stands for to the satisfiability problem in propositional formulas [11]. A SAT solver is a program that determines whether a given SAT instance be satisfied. SMT (Satisfiable Modulo Theories) [11] deals with predicate logics including addition and subtraction on numbers, in addition to propositional logic dealt with SAT. A SMT solver is a program that determines whether a given SMT instance be satisfied.

SAW supports ABC [13], Boolector [7], CVC4 [6], mathSAT [8], Yices [5], and Z3 [4]. Even for the same program, the execution time of the script changes significantly depending on the solver. Therefore, it is necessary to select an appropriate solver according to the algorithm of the program to be verified.

3 PROPOSED METHOD

3.1 Bounded Model Checking

Consider functions that deal with linear list structures. If verification is performed without defining the end of the linear list structure as shown in Figure 2, problems such as failure to complete the verification without completing the recursion and state explosion will occur.

In this paper, in order to solve this problem, we verify the data of the structure using bounded verification method. The bounded verification method verifies in general that the verification property is satisfied for the state obtained by all state transitions from the initial state to a certain given number n, by assumption that execution of a program is regarded as a state transition. In general, bounded verification methods, are realized by performing verification limited number application of unfolding technique to iterations such as loops. In this paper, we define n the size of data structure elements to be given to the function to be verified, in order to realize bounded verification in SAW verification as shown in Figure 3.

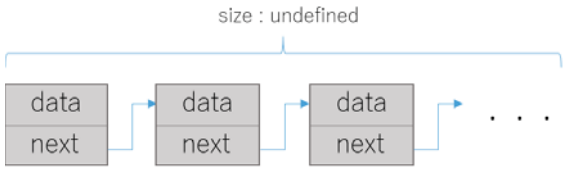


Figure 2: Liner List with Unspecified Size

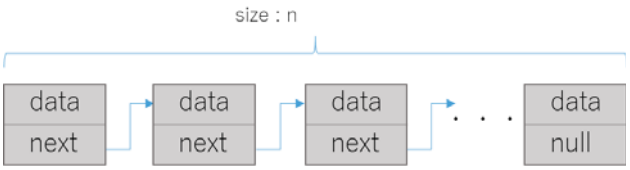


Figure 3: Liner List with Specified Size

3.2 Introduction of Helper Functions to SAWScript

Although it is common to use class variables as function arguments, it is difficult to use class variables as arguments of a function in SAW Script mainly due to the limitation of SAW. Therefore, in this paper, a helper function is used to SAW Script. The behavior of the function under the verification is indirectly verified by verifying the behavior of the helper function. Listing 5 is an example of the helper function. Function `mulfunction` in Listing 5 implements the behavior of Listing 2 in a more complicated way. Although `mulfunction` receives an integer and its instance as its arguments, it is difficult to describe it as a verification property in SAW Script. A helper `helperfunc` is, thus, used.

Listing 6 shows a script that verifies `helperfunc`. As the argument of `helperfunc` is integer, it is possible to perform verification. Moreover, it simplifies the description and have ability to apply wide range of verification target.

In verification, it can be verified indirectly that the behavior of the function `mulfunction` matches the verification property by proving that the function `helperfunc` always returns two times `int` given as an argument.

Listing 5: HelperExample.java

```
1  public class  HelperExample{
2    int tmp = 2;
3
4    int mulfunction(int x, HelperExample test){
5      return x * test.tmp;
6    }
7
8    int helperfunc(int x){
9      HelperExample test = new HelperExample();
10     return test.mulfunction(x, test);
11   }
12 }
```

Listing 6: SAWScript for helper function

```
1  let function_spec : JavaSetup () = do {
2    x <- java_var "x" java_int;
3    java_return {{x * 2 : [32]}};
4    java_verify_tactic abc;
5  };
6
7  java_mul <- java_load_class "HelperExample";
8  java_verify java_mul "helperfunc" []
       function_spec;
```

4 EXPERIMENTS

The following two research questions were established for conducting the evaluation experiment.

RQ1 With the proposed method, can we use SAW to evaluate the behavior of functions that handle recursive structures using Java?

RQ2 How much difference does the verification time produce depending on the solver used?

In order to investigate the above research questions, we implemented two programs that deal with a linear data list structure and a binary tree structure, which are typical programs that handle recursive data structures. Unlike C, Java, in general, uses classes to implement a data structure.

Due to the limitation of SAW, we use a *single* Java class to implement the data structures.

4.1 Linear List

Consider a program that recursively derives the summation of data of all nodes for a linear list structure with $n + 1$ elements. It uses a dummy data at the top node, as shown in Figure 4.

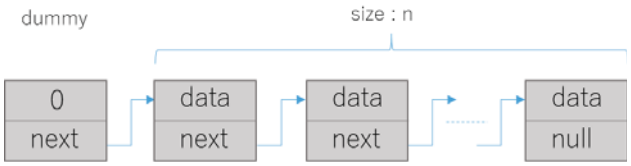


Figure 4: Linear Structure

Listing 7 shows a program that realizes a linear list structure in Java. In order to simplify the description of SAWScript, we described all of them in a single class. Given an int array and a top node of the linear list, the function `insert` adds the node to the end of the linear list and return the top node of the list. If you give the top node to `linearSum`, it recursively derives the summation of the values up to the last node. Also, as a help function for proving, we set up `forProve`.

The function `forProve` firstly generates an instance for an int array as its argument. Next, the function `insert` is called for every element of the array of the arguments. Finally, insertion is performed into the linear list. When insertion ends, the result of calling `linearSum` is returned.

Listing 7: Java Linear class

```
1  public class JavaLinear{
2    int val;
3    JavaLinear next;
4
5    JavaLinear(int val){
6      this.val = val;
7      next = null;
8    }
9
10   int linearSum(JavaLinear now){
11     if(now.next != null){
12       return now.val + linearSum(now.next);
13     }else{
14       return now.val;
15     }
16   }
17
18   void insert(int val, JavaLinear header){
19     JavaLinear now = header;
20
21     while(now.next != null){
22       now = now.next;
23     }
24     now.next = new JavaLinear(val);
25   }
26
27   int forProve(int[] ary){
28     JavaLinear test = new JavaLinear(0);
29     for(int i=0; i<ary.length; i++){
30       insert(ary[i], test);
31     }
32     return linearSum(test);
33   }
34 }
```

Listing 8 is a SAW script for proving Listing 7. The function to be verified is `forProve`. In Line 1 to Line 5, the following items are described. The argument `ary` is an int array with 3 elements; The return value is the summation of all the elements of the argument `ary`; and ABC is used for verification of the property.

Listing 8: SAWScript for Java Linear Program

```
1  let linear_spec : JavaSetup () = do {
2    i <- java_var "ary" (java_array 3 java_int);
3    java_return {{ ((i@0) + (i@1) + (i@2)) }};
4    java_verify_tactic abc;
5  };
6
7  linear_java <- java_load_class "JavaLinear";
8  java_verify linear_java "forProve" []
       linear_spec;
```

4.2 Binary Tree

Consider a program that recursively derives the summation of data of all nodes in a given binary tree structure with $n+1$ elements with dummy data at the root node as shown in Figure 5.

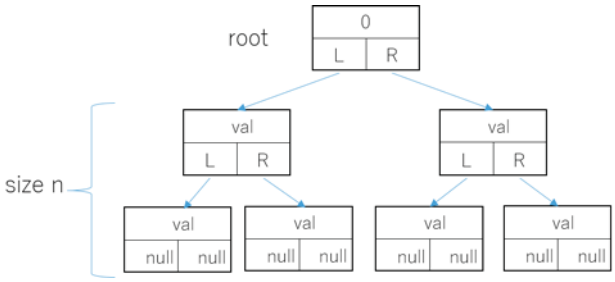


Figure 5: Binary Tree Structure

4.2.1 The First Approach with Insert algorithm

The program that realizes binary tree structure by Java is shown in Listing 9. Also, in this program, all descriptions are given in a single class in order to simplify the description of SAWScript. Given a root node for the function `btreeSum`, it recursively derives the summation of all nodes in the binary tree. We used the function `insert` in Listing 9 as an algorithm to insert nodes into a binary tree. This function receives val and the root of a binary tree as arguments. If $val < now.val$ is satisfied for the current node now, it moves to the left and otherwise right. It finally inserts a new created node for the argument val.

When this algorithm is verified using SAW Script of Listing 10, an error message saying "internal: uncaught unknown exception: Unsupported or mismatched type when merging values: 6::JavaBTree and null)," is generated. Thus, the verification cannot work well for this time and we used alternative approach.

Listing 9: Java Binary Tree Program

```
1  public class JavaBTree{
2    int val;
3    JavaBTree left;
4    JavaBTree right;
5
6    JavaBTree(int val){
7      this.val = val;
8      left = null;
9      right = null;
10   }
```

```
11
12    int btreeSum(JavaBTree now){
13      if(now == null){
14        return 0;
15      }
16      return btreeSum(now.left) + btreeSum(now.
            right) + now.val;
17    }
18
19    void insert(int val, JavaBTree root){
20      JavaBTree now = root;
21      while(true){
22        if(val < now.val){
23          if(now.left != null){
24            now = now.left;
25          }else{
26            now.left = new JavaBTree(val);
27            break;
28          }
29        }else{
30          if(now.right != null){
31            now = now.right;
32          }else{
33            now.right = new JavaBTree(val);
34            break;
35          }
36        }
37      }
38    }
39
40    int forProve(int[] ary){
41      JavaBTree root = new JavaBTree(0);
42
43      for(int i=0; i<ary.length; i++){
44        insert(ary[i], root);
45      }
46      return btreeSum(root);
47    }
48  }
```

Listing 10: SAWScript for Java BTree Program

```
1  let linear_spec : JavaSetup () = do {
2    i <- java_var "ary" (java_array 2 java_int);
3    java_return {{ ((i@0) + (i@1)) }};
4    java_verify_tactic yices;
5  };
6  linear_java <- java_load_class "JavaBTree";
7  java_verify linear_java "forProve" []
       linear_spec;
```

4.2.2 Alternative Approach

For a binary tree with $n = 2$ elements, we created a program to generate 5 patterns of the entire binary tree shown in Figure 6 and replaced it with the function `insert`. The program that defines the binary tree of pattern 1 in Figure 6 is shown in Listing 11 When the verification is performed for the function `forProve_size2_pattern1` by limiting it to the argument $n = 2$, it is confirmed that verification by SAW can be performed. As the behavior of the SAW Script is the same as that shown in Listing 10, we omit it.

As shown above, since the research question in this paper is to use SAW to verify the behavior of a function that handles recursive structures using Java, the behavior of the algorithm for inserting nodes into a binary tree is not the essence of verification.

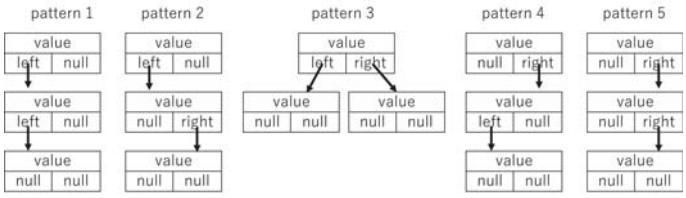


Figure 6: Whole patterns for Binary Tree with size$n = 2$

Listing 11: Alternative Java Program for Pattern 1

```
1  int forProve_size2_pattern1(int[] ary){
2    JavaBTree node1 = new JavaBTree(0);
3    JavaBTree node2 = new JavaBTree(ary[0]);
4    JavaBTree node3 = new JavaBTree(ary[1]);
5    node1.left = node2;
6    node2.left = node3;
7    return btreeSum(node1);
8  }
```

4.2.3 Verification Changing The Number of Elements

As a result of the experiment in the previous section, it was possible to verify the function dealing with the binary tree recursively with SAW. Because the number of elements of the binary tree used in the verification was small, we cannot clarify the performance difference among solvers. Thus, we need to investigate the change of verification time caused by the size of binary tree.

The increase of the patterns of binary tree follows Catalan number, hence, investigating all patterns is unfeasible by time constraint. Therefore, we created complete binary trees in range of the size n is 1 to 14 and verified by SAW. As the behavior of the SAW Script is the same as that shown in Listing 11, we omit it.

5 ENVIRONMENT of THE EXPERIMENTS

The following summarizes the specification of PC used for the experiments.

- PC : TOSHIBA Dynabook T55/76MG
- OS : Windows 10 64bit
- CPU : Intel Core i7 4510U
- RAM : 8GB DDR3

The versions of the tools used in the experiments are as follows.

- SAW : 0.2
- LLVM : 3.8.0
- Java : 1.8.0_211
- ABC : 1.0.1
- CVC4 : 1.5
- Z3 : 4.6.0

- Yices : 2.5.4
- MathSAT : 5.5.1

We use (old) LLVM 3.8.0, due to the SAW limitation.

6 THE RESULTS OF EXPERIMENTS

Here, the experimental results are shown. The figures are the execution time in seconds, rounded down to the fourth decimal place. If the execution time exceeds 600 seconds, it will time out.

6.1 Verification Times of Linear Lists

The results of the verification in Section 4.1 are shown in Table 1.

Table 1: Verification Times for Linear Lists

n	ABC	CVC4	MathSAT	Yices	Z3
1	0.217	0.529	0.475	0.499	0.534
2	0.221	0.495	0.540	0.532	0.507
3	0.429	0.496	0.635	0.533	0.505
4	1.456	0.482	6.843	0.547	0.518
5	6.636	0.486	118.645	0.524	0.476
6	58.352	0.483	450.864	0.544	0.455
7	143.017	0.480	-	0.521	0.506
8	366.254	0.494	-	0.540	0.459
9	481.269	0.485	-	0.491	0.482
10	-	0.558	-	0.485	0.481

6.2 Verification Times of Binary Trees

6.2.1 Verification Times for Alternative Approach

The results of the verification in Section 4.2.2 are shown in Table 2.

Table 2: Verification Times for Binary Trees (A.A.)

pattern	ABC	CVC4	MathSAT	Yices	Z3
1	0.471	0.572	0.624	0.494	0.558
2	0.411	0.597	0.638	0.513	0.596
3	0.388	0.622	0.720	0.491	0.585
4	0.375	0.584	0.683	0.493	0.666
5	0.372	0.594	0.749	0.482	0.588

6.2.2 Verification Times Changing The Number of Elements

The results of the verification shown in Section 4.2.3 are shown Table 3.

Table 3: Verification Times for Binary Trees (C.E.)

n	ABC	CVC4	MathSAT	Yices	Z3
1	0.225	0.595	0.680	0.732	0.773
2	0.137	0.615	0.625	0.783	0.823
3	0.367	0.625	0.882	0.795	0.855
4	3.441	0.650	4.445	0.756	0.742
5	58.232	0.645	118.235	0.804	0.874
6	151.325	0.632	-	0.789	0.830
7	-	0.871	-	0.790	0.884
8	-	0.718	-	0.816	0.852
9	-	0.789	-	0.825	0.900
10	-	0.946	-	0.857	0.821
11	-	0.685	-	0.832	0.834
12	-	0.681	-	0.757	0.943
13	-	0.659	-	0.852	0.824
14	-	0.700	-	0.841	0.876

7 DISCUSSION

7.1 Linear Lists

According to the results of this experiment, verification was possible for all the solvers used. While CVC4, Yices, and Z3 completed verification within one second at $n = 10$, timeout occurred in ABC and MathSAT.

In the previous researches [15], verification results in Table 4 was obtained as a result of verification by SAW with 10 elements for a function that calculates the sum of linear list in C.

Table 4: Verification times for Linear List in C

n	ABC	CVC4	MathSAT	Yices	Z3
10	6.981	1.132	291.854	307.307	59.110

In the verification for Java, we used an integer array as an input, created an instance and inserted it at the end of the linear list. On the other hand, in the verification for C language, we directly described the value of each element of the structure and the next node in SAWScript. One of the differences in verification time that occurred between C and Java would be caused by the difference between the style the SAWScript.

7.2 Binary Trees

7.2.1 Verification Error and Alternative Approach

When Listing 9 is verified, an error message saying "internal: uncaught unknown exception: Unsupported or mismatched type when merging values: 6::JavaBTree and null)," is generated. This error was not occurred when we commented out the lines 26 and 33. Thus, the error is caused by these lines. However, description like this also exists in Listing 7 and no errors were occurred when we verified it.

When we used SAW 0.3, an error message saying "resolveSAWTerm: unimplemented sequence type," is generated. Ac-

cording to this message, SAW doesn't verify this program now. In conclusion, we couldn't resolve this problem and took alternative approach shown in Section 4.2.2.

In the alternative approach, we created helper functions like shown in Listing 11. We verified that it returns the sum of int[] array when it received it. In the helper functions, it defines a binary tree based on its argument. The root node of the binary tree is defined as dummy data and given for function `btreeSum`.

Thus, when the helper function satisfies the verification property, it can be verified inductively that when the function `btreeSum` receive the root node of the binary tree that we defined, it returns the sum of the binary tree.

7.2.2 Verification Time

Generally, the verification time changes by the SMT expression used and solver's spec storingly[14]. It is known as disadvantage of SMT verification and to overcome this problem, we usually use methods shown below.

- Modify the SMT expression.
- Change the SMT solver.

However, SAW creates an SMT expression automatically, users cannot modify it. Thus, we need a method to find a SMT solver that can verify in a short time.

In this paper, we restricted the size of binary tree to n and created helper functions that defines the all pattern of structure of binary tree below size n like Listing 11. Helper function in Listing 11 satisfies that when it receive an array with size n = 2, it returns the sum of the all array by results of SAW verify. In Listing 11, binary trees generated based on the argument and root node is received by function `btreeSum`. Thus, function `btreeSum` satisfies that when it receive the root node of binary tree, it returns the sum of the all nodes in range of $0 \leq n \leq 2$.

When we don't consider nodes' order, the relation of the number of all nodes $x = n + 1$ and the number of patterns of binary tree $P(x)$ is defined in (2) based on Catalan number.

$$P(x) = \frac{2x!}{(x+1)!x!}(x \geq 1) \tag{2}$$

When we consider nodes' order, in other words, when we consider the binary trees in Figure 7 are different, $P(x)$ is defined in (3).

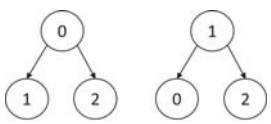


Figure 7: Two binary trees in different orders

$$P(x) = \frac{2x!}{(x+1)!x!}x!(x \geq 1) \tag{3}$$

In this paper, additional verifications were performed with SAW for all binary tree patterns considering the nodes' order in range $1 \leq x \leq 3$ and confirmed all patterns satisfies the verification properties. Thus, the function `btreeSum` returns sum of the btree when it received the root node of the binary tree in range $0 \leq n \leq 2$.

To automate verification the Java program of binary tree, we need to create a program that generates Java functions that define all the binary tree structures exhaustively in range less than any size n. However, the number of binary tree patterns increases based on (3) and it takes amount of time to verify.

Currently, it is difficult to predict verification time by SMT solver[14]. We can find the solver that can verify in a short time by changing the size of data structures.

7.3 Equivalence between C and Java

Though Listing 3 can verify the equivalence of functions written in C and Java, it is due to the simple structure of the functions that have the same input. We have to provide a method to verify the equivalence of functions strictly written in C and Java.

As mentioned before, Java deals with structures in a different way of C. In this paper, helper functions are used for verification of functions using recursive structures described using Java classes. Also, we use `java_verify` in the function verification script.

In the previous research [15], in order to verify the function using a recursive structure described in C, the value held in each element of the structure and the pointer to the next address is described explicitly in the SAW script. The head (or root) of the structures is also given explicitly described. Using these items as arguments, verification is performed successfully. `Crucible_llvm_verify` is used as the verification strategy function in the verification script.

If we have to show the equivalence of multiple functions written in the same language, we can prove the equivalence by showing that the same verification property is satisfied for each pair of functions. However, C and Java is verified in a different manner, and also `java_verify` and `crucible_llvm_verify` are different verification packages.

8 CONCLUSION

We applied our proposed method to Java programs dealing with two types of recursive data structures and verified using SAW, and show that verification is actually possible. For the verification for the linear lists, comparison with the verification for C in the previous research was discussed.

In this paper, we proposed the method to simplify the description of SAW Script and to verify inductive by creating a helper function and defining a structure exhaustively in it. We confirmed that equivalence verification by SAW can be performed in the sense that two functions written in C and Java satisfy the same verification property.

As future work, we will conduct further evaluation experiments and devise a method to verify the equivalence of the behavior of functions dealing with recursive data structures described in C and Java. In addition, we would like to investigate the influence of language and algorithm on the solvers,

and consider a method to find the optimal solver in equivalence verification.

ACKNOWLEDGEMENT

Funding from Mitsubishi Electric Corp. is gratefully acknowledged.

The research is being partially conducted as Grant-in- Aid for Scientific Research C (16K00094), C (17K00111) and A (19H01102).

REFERENCES

[1] E. Clarke, A. Biere, R. Raimi, Y. Zhu: "Bounded Model Checking Using Satisfiability Solving," Formal methods in system design, Vol.19 Issue 1, pp.7-34 (2012)

[2] T. Liu, M. Nagel, and M. Taghdiri, "Bounded Program Verification using an SMT Solver: A Case Study," Proceedings of the 5th International Conference on Software Testing, Verification and Validation, pp.101-110 (2012)

[3] R. Dockins, A. Foltzer, J. Hendrix, B. Huffman, D. McNamee, and A.Tomb: "Constructing Semantic Models of Programs with the Software Analysis Workbench," Proceedings of VSTTE 2016 (2016)

[4] L. de Moura and N. Bjørner: "Z3: An efficient SMT solver," Proceedings of TACAS 2008, LNCS Vol. 4963, pp.337-340 (2008)

[5] B. Dutertre: "Yices 2.2," Proceedings of CAV2014, LNCS Vol.8559, pp.737-744 (2014)

[6] C. Barrett, C. L. Conway, M. Deters, L. Hadarean, D. Jovanović, T. King, A. Reynolds, and C. Tinelli: "CVC4," Proceedings of the 23rd international conference on Computer aided verification (CAV'11) pp. 171-177 (2011)

[7] A. Niemetz, M. Preiner, and A. Biere: "Boolector 2.0," Journal on Satisfiability, Boolean Modeling and Computation, 9, pp.53-58 (2015)

[8] A. Cimatti, A. Griggio, B. Schaafsma, and R. Sebastiani: "The MathSAT5 SMT Solver," Proceedings of TACAS 2013, LNCS Vol. 7795, pp.93-107 (2013)

[9] F. Corzilius, G. Kremer, S. Junges, S. Schupp, and E. Abraham: "SMT-RAT: An Open Source C++ Toolbox for Strategic and Parallel SMT Solving," Proceedings of the International Conference on Theory and Applications of Satisfiability Testing (SAT 2015) pp.360-368 (2015)

[10] N. Een, A. Mishchenko, and N. Sörensson: "Applying Logic Synthesis for Speeding Up SAT," Proceedings of the 10th International Conference on Theory and applications of satisfiability testing pp. 272-286 (2007)

[11] A. Biere, M. Heule, H. Van Maaren, and T. Walsh: "Handbook of Satisfiability," IOS press (2009)

[12] A. Darringer, W. H. Joyner, Jr., C. L. Berman, and L. Trevillyan: "Logic synthesis through local transformations," IBM Journal of Research and Development, Vol.25 (4), pp.272-280 (1981)

[13] R. Brayton and A. Mishchenko: "ABC: An Academic Industrial-Strength Verification Tool," LNCS Vol.6174, pp.24-40 (2010)

[14] Yasunao Takano, Hiroki Sakaji , Shigeyuki Sato: "Toward Automatic Selection and Automatic Tuning of SMT Solver Using Machine Learning," 35th Japan Society for Software Science and Technology (2018)(In Japanese)

[15] R. Karashima, S. Harauchi, K. Okano and S. Ogata: "Proposal and Evaluation for equivalence checking for program with recursive data using SAW," the 25th Workshop of Fundamentals of Software Engineering, pp.91-96 (2018) (In Japanese)

Time Series Segmentation with Leg Analysis for Human Motion Analysis

Makoto Imamura*, Mao Inoue*, Masahiro Terada*, and Daniel Nikovski**

*School of Information and Telecommunication Engineering, Tokai University, Japan
**Mitsubishi Electric Research Laboratories, USA
imamura@tsc.u-tokai.ac.jp

Abstract – To enable automated analysis of human motion data collected by acceleration sensors, gyro sensors, or motion capture devices, an approach for accurately segmenting primitive actions is required. Whereas most existing approaches use templates of basic actions such as “stand up”, “walk” and “sit down”, we introduce a novel problem called “structural motif discovery” that aims to find segments without templates from repetitive routine motion that consists of regularly ordered (actions?). We also propose a novel segmentation method that approximates the time series with a sequence of convex-shaped patterns by means of leg analysis, which is parameter-free and its complexity is $O(N)$, where N is the length of a given time series. The experimental results show that our proposed method is effective for both simulation data and real data from repetitive assembly operations.

Keywords: Time series, Segmentation, Human motion analysis, Sensor data mining, Convex-shaped pattern

1 INTRODUCTION

To enable automated analysis of human motion data such as exercise monitoring, gesture recognition, human machine interaction, and robot imitation learning, segmenting primitive actions is critical [1][2][3]. Time series segmentation is the process of identifying the temporal events of movements of interest, making a continuous sequence of time series into smaller subsequences to facilitate movement identification, modeling, and learning.

We have developed two applications of human motion analysis with acceleration sensors, gyro sensors, and motion capture devices. The application domains are factory work processes [4] and baggage lifting work [5]. In these applications, time series segmentation is critical. Therefore, we propose a novel segmentation method that approximates the time series with a sequence of convex-shaped patterns from a continuous motion sequence with leg analysis [6].

The rest of our paper is organized as follows. Section 2 describes the problem statement, our approach, and the scope of this paper. Section 3 describes our method that consists of convex feature extraction and symbolic convex approximation. Section 4 evaluates our method on one simulation and one real data sets.

2 BACKGROUND

2.1 Problem Statement

A typical method for human motion analysis is a combination of template matching by dynamic time warping (DTW) and change point detection by segmentation according to Zero Velocity Crossing (ZVC) with given general templates such as “stand up”, “walk” and “sit down” [1][2][3]. However, factory worker motions are complicated and depend on the target product and the target process, so it is difficult to prepare specific templates that correspond to such motions for the purpose of analysis. Therefore, we need to deal with the problem of how to extract basic actions without templates from time series collected by sensors such as acceleration, gyro and motion capture. The type of motion targeted in this paper is a repetitive routine operation that consists of regularly ordered basic actions.

Figure 1 shows an example of a repetitive routine operation. The routine operation consists of three basic actions, which are “(a) carry a main body product from a previous process”, “(b) attach a part to the main body product (action B)” and “(c) carry the main body product to the next process”. Hereafter, we call a repetitive routine process a *cycle*. A cycle is important in the factory domain, because it corresponds to a process in production. If we can extract a cycle correctly, we can measure the working time of a process. The automatic measurement of each working process time for each worker helps us to find bottleneck operations, an operation error, and also possibly measure the fatigue level of the worker, which are useful for improving product efficiency.

The cycle in Figure 1 is expressed by ”abbc”, ”abbbc” or “abc”, if the action “b” repeats more than once depending on the specification of a product. By using a regular grammar, the cycle can be expressed by ”a(b+)c”. The problem in this paper is described by the following.

Problem: Let X be a time series that corresponding to a series of repetitive routine operations that consist of regularly ordered actions. Find a repetitive routine pattern in X and find a regular expression to represent that pattern.

Hereafter, we call the above problem *structural motif discovery*.

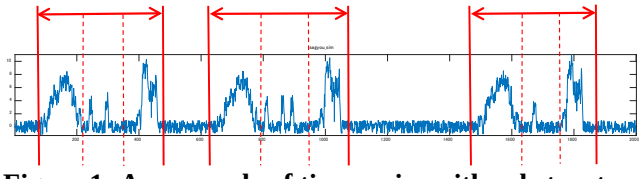


Figure 1: An example of time series with substructure

2.2 Approach

A typical existing method to solve the structural motif discovery problem is motif discovery [7][8]. Because the subsequences corresponding to the same action are similar to each other, we can find the basic actions as repeatedly occurring subsequences by means of motif discovery. However, existing motif discovery algorithms require subsequence lengths for basic actions. In our problem setting, we do not know the subsequence length of a basic action. Furthermore, the complexity order of motif discovery is $O(N^2)$ or $O(N^3)$, where N denotes the length of the time series, because of the embedded calculation of Euclidean or DTW distances.

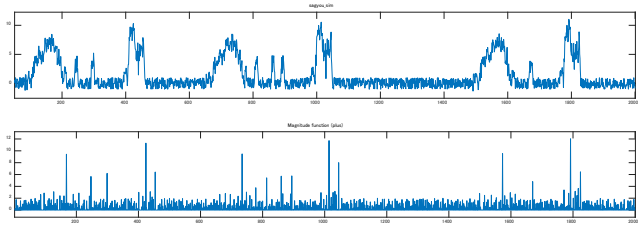


Figure 2: An example of the result by leg analysis

We investigate a novel approach to approximating a time series with a series of convex-shaped patterns by means of a leg analysis algorithm, which is parameter free and its complexity order is $O(N)$. We expect that our approach has advantages over the existing motif discovery algorithms with Euclidian or DTW distances in the following aspects.

(i) Feature extraction: The highest-cost process in solving structural motif discovery problems is the feature extraction to recognize actions. The feature extraction in our approach is convex-shaped pattern extraction by leg analysis, whereas that in existing approaches is nearest neighbor search with Euclidian distance. Our computational complexity is O(N), vs. $O(N^2)$ of existing methods.

(ii) Parameter selection: A required parameter of our approach is the magnitude level of a convex-shaped pattern, whereasexisting methods require the subsequence length. Though the subsequence length varies for each basic action, the magnitude level can be expected to be more independent from each basic action, because the magnitude level depends on small primitive human motions that construct a basic action. This hypothesis is perhaps not self-evident, so it requires exhaustive experiments to test it.

(iii) Parameter optimization: When we try another parameter value, our approach does not need recalculation for feature extraction, which existing methods do. The difference is that leg analysis is parameter-free so that extracted features are also independent of this parameter. Our approach is easier to try for various parameter values than existing methods.

(iv) Robustness: Our approach approximates time series with convex-shaped patterns, while existing methods treat a strict shape by Euclidean distance. This difference has pros and cons. With regard to robustness to the noise in time series, our approach has the possibility to be superior. Structured motif discovery requires a huge combinatorial calculation cost, therefore an approximation approach has the possibility to be able to prune search trees efficiently. On the other hand, with regard to the precision of pattern recognition, existing methods might be superior to our method. This hypothesis also requires exhaustive experiments to test it.

2.3 Scope

The purpose of this paper is an early stage trial to confirm if our approach can provide an insight into the structural motif discovery problem. We formulate our approach and evaluate it one two examples: one simulation data set and one real data set. A quantitative evaluation on exhaustive experiments and parameter optimization methods are left as future work.

3 METHOD

3.1 Convex Feature Extraction

Leg analysis [6] provides a convex pattern extraction algorithm to extract every convex-shaped pattern which has local maximal or local minimum values. A convex-shaped pattern is characterized by a 4-tuple (“peak position”, “magnitude”, “left terminal”, “right-terminal”). In a precise definition, we should treat a trapezoid as a convex-shaped pattern in the case when its peak is a flat region, but here we only show a simplified definition of it for readers to understand the key idea of our approach. The precise definition and a detail algorithm for its computation are covered by the original paper [6].

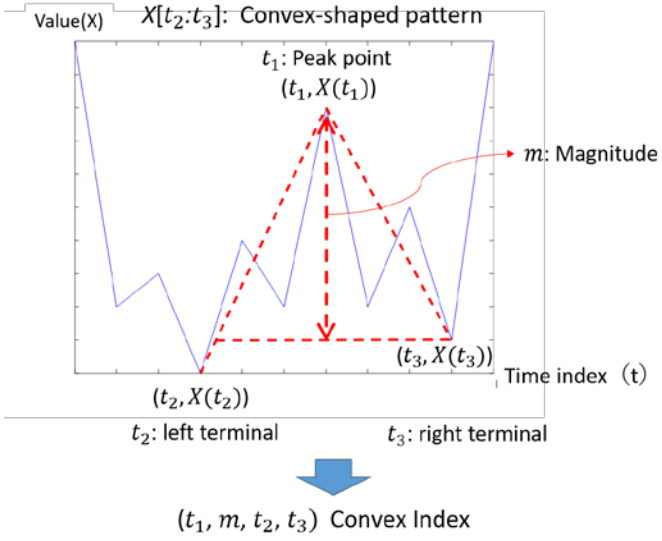


Figure 3: An example of a convex-shaped pattern and its convex index

We define a time series, a subsequence, a convex pattern index and a convex profile as a data-structures to describe a convex-shaped pattern. Figure 3 shows an example of a convex-shaped pattern and its convex index.

Definition: *Time series*

A *Time Series* $X=[x_1,\cdots,x_m]$ is a continuous sequence of real values. The value of the i-th time index is denoted by $X[i] = x_i$.

Definition: *Subsequence*

A *subsequence* $S = [x_p, x_{p+1},...,x_q] = X[p:q]$ is a continuous subsequence of X starting at position p and ending at position q. We denote the length of a subsequence S by len:

$$len(S) \equiv q - p + 1$$

Definition: *Convex index*

A convex index is a 4-tuple (peak position, magnitude, left position, right position). A peak position is a time index at which the peak of a convex-shaped pattern is. A magnitude is a height of a convex-shaped pattern. A positive magnitude value means that a pattern is convex, and a negative magnitude value means that a pattern is concave. A left terminal is a time index at which a convex-shaped pattern starts. A right terminal is a time index at which a convex-shaped pattern ends.

Definition: *Convex profile*

Let X be a time series. A convex profile of X is a list of convex indexes that correspond to an output of convex pattern extraction process for input X. We denote a convex profile as CP and the i-th convex index in CP as CP(i). We denote a peak position, magnitude, left position and right position of CP(i) as CP(i).p, CP(i).m, CP(i).l and CP(i).r respectively.

Convex pattern extraction is represented as a function from a time series X to a convex profile of X. A subsequence corresponding to a convex index C(i) is represented as X[C(i).left: C(i).right] where X is a time series. We also define a magnitude function to plot the magnitude values to visualize the outline of a convex profile.

Definition: *Magnitude function*

Let X and CP be a time series and the convex profile of X respectively. Magnitude function *mf* is a function from each time index of X to a real value.

If t is a time index at which a convex index CP(i) has a value as a peak position, then *mf*[t] =CP(i).m, otherwise *mf*[t] = 0.

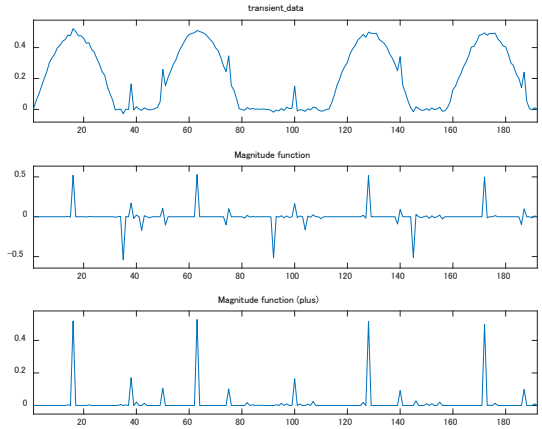


Figure 4: An example of a magnitude function

Figure 4 shows an example of a time series and its magnitude functions. Top graph is a line graph of the original time series data. The second graph is its magnitude function. We see that a convex-shaped pattern in the original data corresponds to a spike in the magnitude function. The bottom graph is the magnitude function that has only positive values. Heareafter, we use only positive magnitudes in this paper.

Table 1 shows an example of a convex profile of a time series shown in Figure 3. One line of a convex profile corresponds to one peak value in the magnitude function.

Table 1: An example of a convex profile

Peak position	**Magnitude**	**Left** termnal	**Right** terminal
16	0.52	1	32
38	0.17	35	42
50	0.11	35	51
63	0.53	35	92
75	0.10	74	81
100	0.16	92	104
128	0.52	92	145
140	0.09	139	145
172	0.50	145	190
187	0.10	186	190

3.2 Symbolic Convex Approximation

Symbolic convex approximation is a means to approximate a time series with a set of convex-shaped patterns by means of a magnitude constraint bin, which is a rule for selecting a convex index and its mapping to a symbol. A magnitude range bin is defined by the following.

Definition: *Magnitude constraint bin*

A magnitude constraint is a logical formula that consists of inequalities on magnitude. A magnitude constraint bin is a list of range constraints which are exclusive of each other.

A convex decomposition is a set of convex-shaped patterns to approximate a time series. A required magnitude level depends on the application, and we can control a magnitude constraint bin to get convex-shaped patterns that are suitable for an application. Symbolic convex approximation is a sequence of symbols used to extract a cycle that consists of regularly ordered patterns by means of a string matching technique. A convex decomposition and a symbolic convex approximation are defined in the following.

Definition: *Convex decomposition*

Let CP be a given convex profile, and let MB be a magnitude constraint bin. A convex decomposition is a set of convex patterns in CP that satisfy either one of the magnitude constraints in MB. Each convex index $C(i)$ is modified to $C'(i)$ so as not to cross another convex pattern by the following formulas:

$$C'(i).l = \max(C(i).l, \max\{C(j).r \mid j \in \{j|C(j).r < C(i).p\}\})$$

$$C'(i).r = \min(C(i).r, \min\{C(j).l \mid j \in \{j|C(j).l > C(i).p\}\})$$

Definition: *Symbolic convex approximation*
Let CP and MB be a convex profile and a magnitude constraint bin respectively. If a convex index C(*i*) satisfies a magnitude constraint *j*, symbolic convex mapping is defined to be a function from C(*i*) to a symbol that corresponds to *j*. That is, if a magnitude constraint *j* is different, the mapped symbol of *j* is different. The image of a symbolic convex mapping from a convex profile is called a symbolic convex approximation.

Figure 5 shows the convex decomposition and the symbolic convex approximation of a raw data shown in the top graph of Figure 4. The top graph is the line graph for the original data. The second graph is a convex decomposition that corresponds to a magnitude constraint "M >= 0.4", which means its magnitude is greater than or equal to 0.4. The corresponding convex shaped-patterns are labeled as "A". The third graph is a convex decomposition that corresponds to a magnitude constrant "0.075 < M and M<=0.4". The corresponding convex shaped-patterns are labeled as "B". The bottom graph is a convex decomposition that corresponds to a magnitude constraint bin {"M >= 0.4", "0.075 < M <=0.4"}. Note that the left and right terminals of the extracted convex-shaped patterns are modified by the definition of convex decompostion. The symbolic convex approximation by this decomposition is "ABBABBABAB".

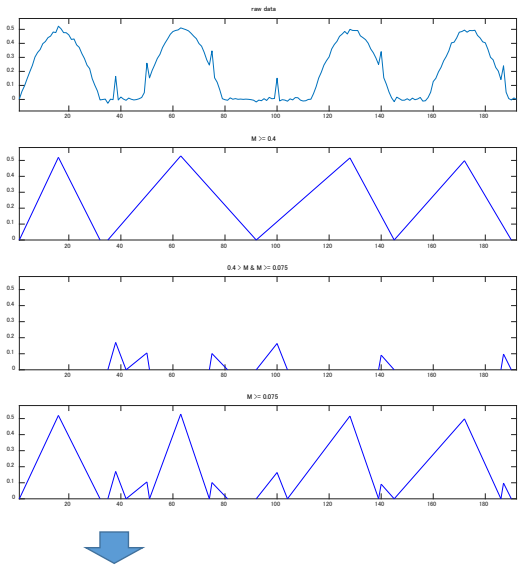


Figure 5: An example of a convex decomposition and its corresponding symbolic convex approximation

4 EXPERIMENTAL VERIFICATION

This section evaluates our method on one simulation and one real data sets.

4.1 An experiment on a simulated time series

Figure 6 shows the magnitude functions corresponding to serveral magnitude constraints for the simulated time series shown in Figure 1. The top graph is a line graph of the simulated data. The second graph is a magnitude function graph that corresponds to a magnitude constraint "M > 0", which means the magnitude value is greater than 0. The third graph is the magnitude function that corresponds to a magnitude constrant "8.5 < M". This graph shows that the convex patterns "a" and "c" in Figure 1 can be extracted by this magnitude constraint. The bottom graph is the magnitude function that corresponds to a magnitude constrant "4.5 < M and M<=8.5". This bottom graph shows that a magnitude convex pattern "b" in Figure 1 can be extracted by that magnitude constraint.

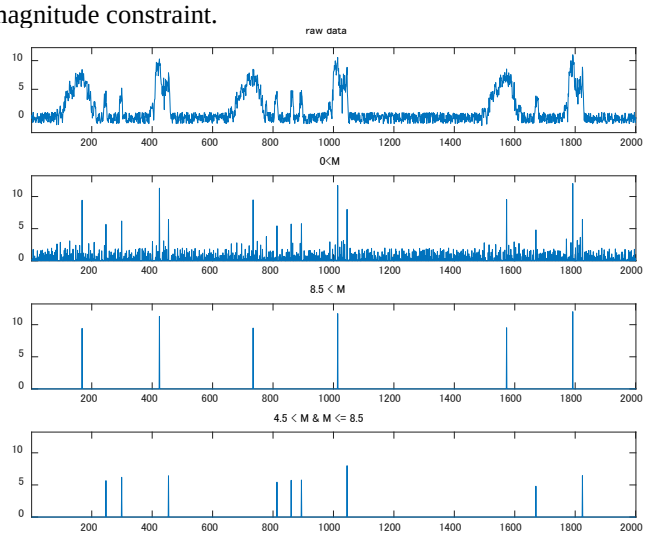


Figure 6: The magnitude functions of a simulated data set

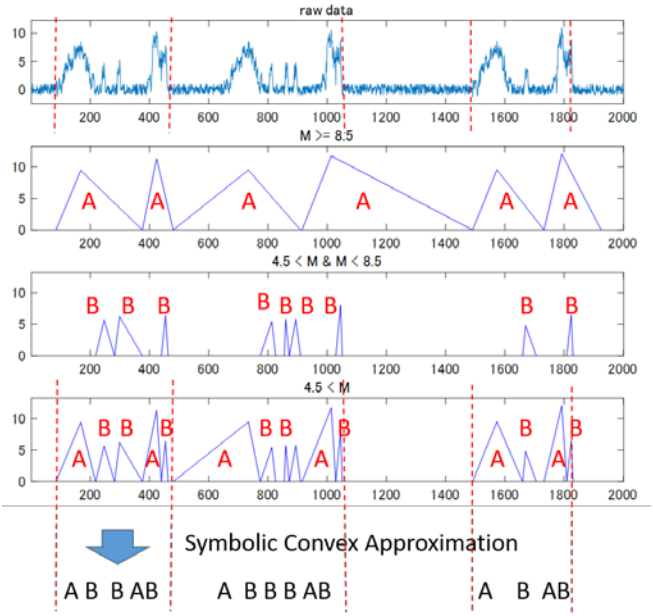


Figure 7: The convex decompositions and the symbolic convex approximation of simulated data

Figure 7 shows the convex decomposition and the symbolic approximation of the simulated data. The top graph is the line graph for the original data. The second graph is a convex decomposition that corresponds to a magnitude constraint of "M >= 8.5". The extracted convex-shaped patterns are

labeled as “A”. The third graph is a convex decomposition that corresponds to a magnitude constrant “4.5 < M and M<=8.5”. The extracted convex-shaped patterns are labeled as “B”. The bottom graph is a convex decomposition that corresponds to a magnitude constraint bin of {“M >= 8.5”, “4.5 < M and M<=8.5”}.

The symbolic convex approximation by this decomposition is “ABBABABBBABABAB”. If we replace a pattern “AB” followed by the repetition of pattern “B” with “c”, we get “ABBcABBBcABc”. After this step, if we replace “A” with “a” and replace the repetition of “B” with (b+), we get “a(b+)ca(b+)ca(b+)c”. This regular exparesion is what we wanted to find. The above procedure contains an ambiguous step, because a single occurrence of “B” after “A” can match two patterns “AB” and A(B+). In the case of this example, we can disambiguate it to use the peak positions of convex shaped patterns “A” and “B”. An algorithm to aquire regular expression by using both a magnitude and a peak position can be considered in future work.

4.2 Factory Work Process

The second example involves human motion in a factory work process. Figure 8 shows an acceleration time series for a cardboard packaging process. Each cycle consists of 4 basic operations: (a) Preparing a cardboard, (b) labeling, (c) packaging and (d) carrying a cardboard. The repeated process can be segmented by discovering a motif, which is a subsequence frequently occurring in a time series. In Fig. 1, the subsequences corresponding to (a), (b), (c) and (d) are examples of motifs.

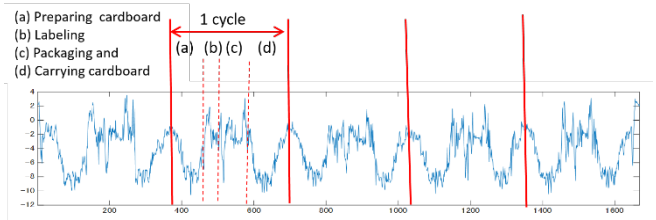


Figure 8: Time series of factory working process

The segmentation problem for this data is to extract each cycle that consists of basic actions (a), (b), (c) and (d).

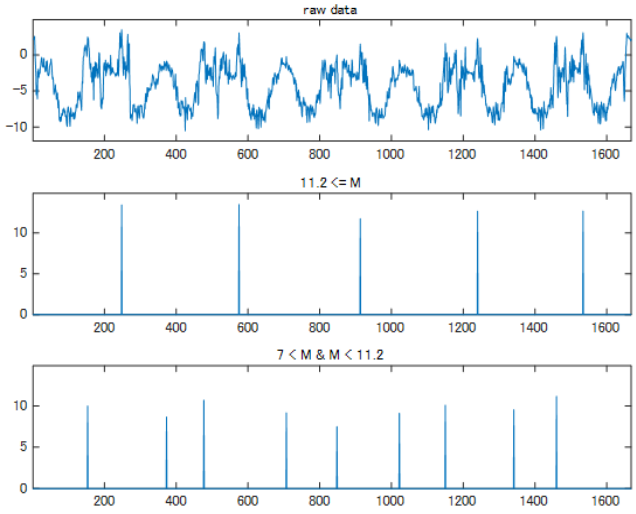


Figure 9: The magnitude function of a factory work process data set

Figure 9 shows the magnitude functions corresponding to serveral magnitude constraints for a factory work process data set shown in Figure 8. The top graph is the line graph for the raw data. The second graph is the magnitude function that corresponds to the magnitude constraint “0 < M”. The third graph is the magnitude function that corresponds to the magnitude constrant “11.2 <= M”. This graph shows that the peak of a segment “c” in Figure 8. can be extracted by this magnitude constraint. The bottom graph is the magnitude function that corresponds to a magnitude constrant “7 < M < 11.2”. This bottom graph shows that the starting point of a segment “a” and the peak of a segment “b” in Figure 8 can be extracted by that magnitude constraint.

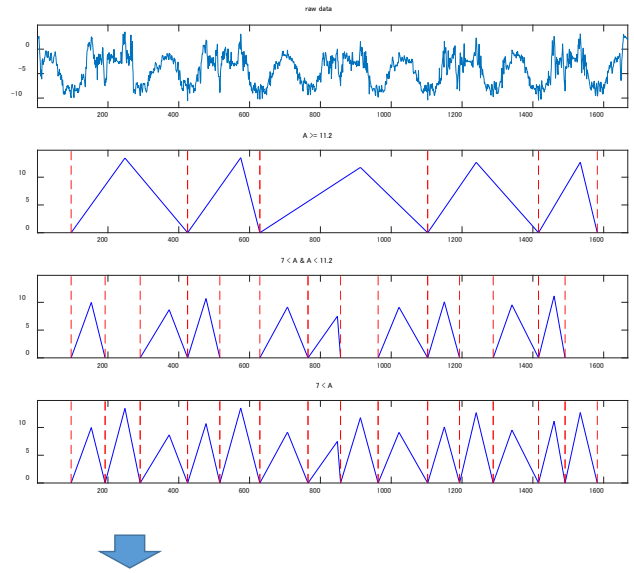


Figure 10: The convex decomposition and the symbolic convex approximation of a factory work process data set

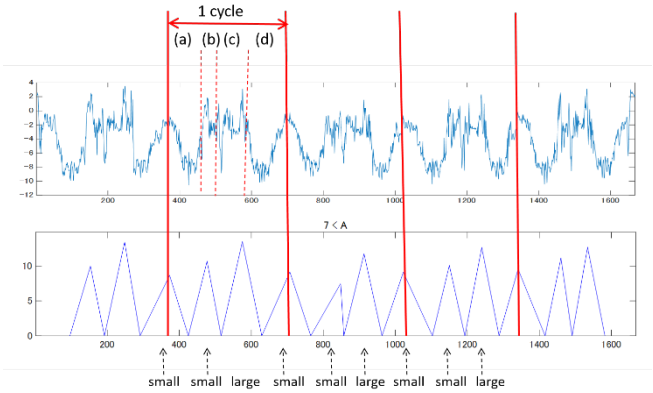


Figure 11: The relation between the segments in a factory work process data set and the extracted convex-shaped pattern

Figure 10 shows the convex decompositions and the symbolic approximation of the factory work process data. The top graph is the line graph for the original data. The second graph is the convex decomposition that corresponds to a magnitude constraint "11.2 <= M". The extracted convex–shaped patterns are labeled as "A". The third graph is the convex decomposition that corresponds to the magnitude constrant "7 < M <11.2". The extracted convex-shaped patterns are labeled as "B". The bottom graph is a convex decomposition that corresponds to the magnitude constraint bin {"11.2 <= M", "7 < M <11.2"}.

The symbolic convex approximation by this decomposition is "BABBABBABBABAB". This symbolic sequence does not correspond to the exact segments shown in Figure 8. However, a repeated sequence "BAB" corresponds to one cylcle, so that we can calculate the working time per one cycle. This result is useful from the point of view of our application. Figure 11 shows the relation between the segments in the factory work process and the convex-shaped pattern extracted by our algorithm. The peak of convex-shaped pattern "A" seems to correspond to the starting point of a segment "a" and the ending point of a segment "d". How to obtain segments required by a given application by the convex-shaped pattern will also be considered in future work.

5 CONCLUSION

We defined the structural motif discovery problem and proposed a novel segmentation method that approximates the time series with a sequence of convex-shaped patterns by means of leg analysis. The experimental results show that our method has the potential to solve the structural motif discovery problem. We only evaluated our method on two simple examples. Evaluation on exhaustive data is future work. Technical future challenges include the following:

(1) Magnitude constraint selection

We manually tuned magnitude constrains depending on the experimental data. How to select an appropriate magnitude constraint is the first major challenge.

(2) Regular expression acquisition

An algorithm to aquire regular expression by using both magnitude and a peak position is also future work.

(3) Segmentation with a convex-shaped pattern

How to obtain segments required by a given application by the convex-shaped pattern will also be considered in our future work.

This work is supported by JSPS KAKENHI Grant Number 17K00161.

REFERENCES

[1] J. F. Lin and D. Kulić: "Online Segmentation of Human Motion for Automated Rehabilitation Exercise Analysis," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 22, no. 1, pp. 168-180, Jan. 2014.

[2] J. F. Lin, M. Karg and D. Kulić: "Movement Primitive Segmentation for Human Motion Modeling: A Framework for Analysis," in *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 3, pp. 325-339, (2016).

[3] M. Reyes Adame, A. Al-Jawadm M. Romanovas, M. A. Hobert, W. Maetzler, K. Möller, Y. Manoli: TUG Test Instrumentation for Parkinson's disease patients using Inertial Sensors and Dynamic Time Warping, Biomed. Eng., vol. 57, pp. 1071-1074, (2012)

[4] Masahiro Terada, Tadao Yago, Mao Inoue, Kakeru Ota, Hiroto Anzai and Makoto Imamura: An analysis Method of Baggage Lifting Motion for lower backache Prevention, IEEJ (The Institute of Electrical Engineering of Japan) 2019 Annual Conference on Electronics, Information and Systems (2019). (in submission)

[5] Tadao Yago, Mao Inoue, Junji Tsuda and Makoto Imamura, Factory Work Process Segmentation Using Acceleration Sensor, Asia Pacific Conference on Robot IoT System Development and Platform 2018 (APRIS2018), (2018).

[6] M. Imamura, J. Tsuda, D. Nikovski, and M. Tsuru : A Fast Online Algorithm for Analyzing Magnitude Fluctuation of Time Series, International Journal of Informatics Society (IJIS), Vol.10, No.2,pp. 85-96 (2018).

[7] Patel, P., Keogh, E., Lin, J. and Lonardi, S.: Mining motifs in massive time series databases. 2002 IEEE International Conference on Data Mining (ICDM), pp. 370-377 (2002)

[8] Chin-Chia Michael Yeh, Yan Zhu, Liudmila Ulanova, Nurjahan Begum, Yifei Ding, Hoang Anh Dau, Diego Furtado Silva, Abdullah Mueen, Eamonn Keogh : Matrix Profile I: All Pairs Similarity Joins for Time Series: A Unifying View that Includes Motifs, Discords and Shapelets, 2016 IEEE International Conference on Data Mining (ICDM), pp. 1317-1322 (2016).

Finding Common Features Among Multiple Groups in Sparse Feature Selections: A Case Study with Wagyu Data

Nanami Higashiguchi[†*], Masatsugu Motohiro[†**],
Haruka Ikegami[§], Tamako Matsuhashi[§], Kazuya Matsumoto[§]
and Takuya Yoshihiro[‡**]
[†]Graduate School of Systems Engineering, Wakayama University, Japan
[§] Graduate School of Biology-Oriented Science and Technology, Kinki University, Japan
[‡]Faculty of Systems Engineering, Wakayama University, Japan
{*higashiguchi.nanami, **motohiro.masatsugu}@g.wakayama-u.jp
*** tac@wakayama-u.ac.jp

Abstract - Wagyu is known in the world as a branded beef of Japan. We are exploring how to predict Wagyu beef quality from protein expression profiles of early-stage beef cattle. Since the protein expression data has a large amount of proteins, we must select a part of them that is truly correlated with beef quality. As the sparse linear regression method, LASSO (Least Absolute Shrinkage and Selection Operator) is the best-known. Although LASSO retrieves a small number of features that explains the target traits, it does not aware groups of samples that has different trends. Unfortunately, it is known that Wagyu data has different trends with each branded region because of the difference in breeding methods of Wagyu beef. In this paper, we propose a method to select features that commonly effects on beef quality among multiple regions as well as features specific to each region. By identifying those two sorts of features separately, we would have better understanding on the elements to improve beef quality, and would be possible to predict beef quality more accurately than normal LASSO regression.

Keywords: Wagyu, Carcass Characteristics, Feature Selaction, LASSO, Sparse Analyses

1 Introduction

Wagyu is known as a high-quality branded beef of Japan, with a feature of soft and tender meats due to fats mixed in the meat. There are several regions famous for Wagyu in Japan, and each region has different policy of breeding sires and beef cattle to produce larger amount of higher-quality meat. This contention among regions has improved the methodology of breeding beef cattle so far. However, since they mostly depends on traditional methods based on statistics on bloodlines or breeding experience, there is a clear limitation of improvement.

Recently, several comprehensive analysis in genomics or proteomics have been developed; for example, gene and protein expression profiles that include expression values of so many genes and proteins are available with smaller cost than ever. Specifically, we have a large number of explanation variables retrieved from each samples, which potentially makes us predict beef quality of each beef cattle in the early stage of beef-cattle breeding. This also could lead to the innovative methodology of breeding beef cattle to improve its beef quality.

Here, the first problem is that the variables in genes or protein profiles are so many that we can hardly select the optimal variable set to predict beef quality. The second problem is that beef cattle of distinct regions has different trend on its data so some analytic methods to treat this problem is required. As for the first problem, recently sparse analyses have been developed in which near optimal feature selection is possible with small computational cost. Especially, if we intend to perform multiple regression, LASSO (Least Absolute Shrinkage and Selection Operator) [1] is often used. LASSO minimizes MSEs (Mean Square Errors) in the form of multiple linear regression, in which by using L1 regularizer most of the coefficients are to shrink to zero. LASSO actually selects a near-optimal variable set within feasible time even if the number of available variables is very large. However, LASSO has a problem in Wagyu analysis that it cannot catch up with the trend of each branded Wagyu regions.

Multi-task LASSO[2], which considers multiple objective functions in selecting a variable set has been proposed. Multi-task LASSO applies L1/L2 penalty to retrieve a variable set that commonly explains the multiple objective functions. This by definition can be used to explain the trend of each region of Wagyu brand by retrieving the common variable that explains trends of the all target regions. However, it is known in both LASSO and Multi-task LASSO that the selected variables are not always optimal in terms of multiple regression so that we can hardly retrieve the optimal set of variables that explains the target traits of Wagyu beef [3]. Methods to retrieve the optimal variable set while considering multiple Wagyu brand regions are required.

In this paper, we present a solution for this problem, i.e., we propose a variable selection method that retrieves an optimal commonly effecting variables among multiple Wagyu regions within a feasible computational time. We exploit a single regression results, i.e., correlation coefficients, and fairness indices among them to retrieve a small number of variables as a candidate of selected variables. By testing all combinations of the candidate variables in multiple regression, we finally retrieve the best variable set within feasible time.

This paper is organized as follows. In Section 2, we describe the trend of Branded Wagyu beef. In Section 3, we introduce LASSO and Multi-task LASSO. In Section 4, we present a proposed method, and evaluation results are shown in Sectio 5. Finally we conclude the work in Section 6.

2 A Brand Wagyu Beef

Japanese Black Cattle is a beef cattle peculiar to Japan, which produces various brand beef called Wagyu such as Kobe beef, etc. There are many regional brand beefs in Japan, each of which has its own way to breed cattle, and apply its own criterion to authorize whether each head of cattle is sold under the name of the brand beef. As the authorization criteria, there are several items, e.g., the birth of cattle, the way to raise cattle, the rating of beef, etc. Among them, the rating of beef is the most important. The rating criteria include various values, and especially 6 items among them are regarded as the most important ones to judge whether a head of cattle is authorized as brand beef [4]. The 6 items, which we call *economical traits*, are CW (Carcass Weight), BMS (Beef Marbling Standard), YE (Yield Enhancement), RT (Rib Thickness), SFT (Subcutaneous Fat Thickness), and REA (Rib-Eye Area). Basically from these criteria, the price of beef in the market is determined. Therefore, the farmers of brand beef have been made a great endeavor to produce quality beef.

Wagyu farmers take various methodologies to produce quality beef stably. One of the most important methods is to control bloodline so as to have better values of the economical traits. Since the bloodline is known to have close relationship with economical traits, efficient inbreeding by producing and identifying genetically excellent individual cattle has a significant importance to improve the value of brand beef site. Each brand-beef site usually breeds several head of cattle called sires that have excellent genetic ability [5][6]. From sires, we take sperms and freeze them, and sell them to farmers. With this system, excellent bloodline of sires is distributed to farmers and generate thousands of children cattle from an excellent sire. Note that, in brand-beef sites, father of each beef cattle is called '1-generation ancestor' and the father of beef cattle is one of the most important criteria to predict economical traits of beef cattle.

As a statistical methodology to predict economical traits of beef cattle from past records, the breeding values are usually used in brand beef sites. The breeding values are calculated for each economical trait, which represent the ability to improve the trait values compared to the average ability in the group. There are two kinds of breeding values, i.e., estimated breeding values and expected breeding values. The former is calculated for sires who have descendants with carcass characteristic scores and represents the ability to improve 6 economical traits. In contrast, the latter is calculated for each beef cattle that does not have enough number of descendants to estimate breeding values.

We have several variations of bloodline models used to compute breeding values. Currently, the most frequently used model is so called 'animal model,' which considers all the relative relationship including brothers of beef cattle that have the same mother. With a bloodline model and the data set, BLUP method calculates the breeding values in a statistical manner as the genetic ability inherited through bloodlines [7]. The expected breeding value for each beef cattle is calculated as the average of its two parents.

On the other side, raising method to produce high-value beef cattle stably also has been studied so far. However, methods in this area are mostly depends on experiences of farmers, and are not based on any scientific results or real data. For example, livestock associations or stock farmers have accumulated their experience to raise high-value beef cattle as know-how or some kind of manuals. This kind of information has wide variations from direct methods such as how to feed cattle to indirect methods such as the structure of cowsheds. As for the academic results, a few studies have been published on the relationship between raising methodology and economical traits. For example, there is a study on improving BMS values by controlling the concentration of vitamin A [8]. However, in the current state, we have still too little knowledge to actually control economical traits in raising in livestock farms.

3 LASSO and Multi-task LASSO

LASSO [1] is a well-known technique for feature selection from the large number of features based on linier regression models. Let S be the given set of samples, and F be that of features. Let $x_{sf}(s = 1, 2, \ldots, |S|, f = 1, 2, \ldots, |F|)$ be the measured feature value of sample s on feature f, where $|S|$ and $|F|$ are the number of elements of S and F, respectively. Hereafter we may write just S and F in place of $|S|$ and $|F|$ for conciseness. Let $\mathbf{x}_s = (x_{s1}, x_{s2}, \ldots, x_{sF})^{tr}$, be the measured vector for each sample $s \in S$, where A^{tr} denotes a transposed matrix of A. Let $X = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_S]$ be the matrix of the feature data. Let y_s be the measured trait values for each sample s, and $\mathbf{y} = (y_1, y_2, \ldots, y_S)$ be the trait vector. Then, LASSO is formulated as follows:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}}(\|\mathbf{y} - \boldsymbol{\beta}\mathbf{X}\| + \lambda|\boldsymbol{\beta}|) \tag{1}$$

where λ is a non-negative regularization parameter, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_F)$ is a coefficient vector for X. Additionally, $\|\boldsymbol{\beta}\|$ represents the L2 norm of a vector $\boldsymbol{\beta}$ defined as $\|\boldsymbol{\beta}\| = \sqrt{\sum_{f\in F} \beta_f^2}$, and $|\boldsymbol{\beta}|$ represents the L1 norm defined as $|\boldsymbol{\beta}| = \sum_{f\in F} |\beta_f|$ Due to the effect of L1-norm penalty with λ, most of β_f converges to zero during the computation of the optimal solution. As a result, we have a small number of non-zero coefficients, and this process works as a feature selection from a large number of feature variables.

Multi-task LASSO [2] is an extension of LASSO, which treats multiple objective functions. Let us denote T as the set of tasks (i.e., set of objective functions), and also let us define a feature matrix and a trait vector for each task. Namely, we let $x_{sf}^{(t)}$ be the measured feature values for task $t \in T$. Also, let $y_s^{(t)}$ be the measured trait values for task $t \in T$. Similarly, we also write $\mathbf{x}_f^{(t)}$, $\mathbf{X}^{(t)}$, $\mathbf{y}^{(t)}$, $\boldsymbol{\beta}^{(t)}$ etc. Note that the number of samples for each task $S^{(t)}$ could be different. Then, the multi-task LASSO is expressed as follows:

$$\hat{\boldsymbol{W}} = \arg\min_{\boldsymbol{W}}(\sum_{t\in T} \|\mathbf{y}^{(t)} - \boldsymbol{\beta}^{(t)}\mathbf{X}^{(t)}\| + \lambda|\mathbf{w}|), \tag{2}$$

where W is the coefficient matrix that combines all coefficient vectors, defined as $W = [\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \ldots, \boldsymbol{\beta}^{(T)}]$, and

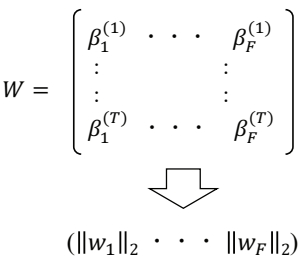


Figure 1: L1/L2 Regularization in Multi-task LASSO

$\mathbf{w}$ is the vector of L2 norms of the coefficients where $\mathbf{w} = (\|\mathbf{w}_1\|, \|\mathbf{w}_2\|, \ldots, \|\mathbf{w}_F\|)$, and $\mathbf{w}_f$ for $f \in F$ is defined as $\mathbf{w}_f = (\beta_f^{(1)}, \beta_f^{(2)}, \ldots, \beta_f^{(T)})$. Note that the regularization term is the combination of L1 and L2 norms, as shown in Fig. 1. In multi-task LASSO, the coefficients $\beta_f^{(t)}$ are defined for each $f \in F$ and $t \in T$. To proceed feature selection and coefficients optimization altogether, multi-task LASSO uses the combination of L1/L2 regularization. First, L2-norm of w_f, the coefficients vector of the same feature is computed, and second, L1-norm of those are used in the regilarization term. This enables us to select the commonly effective features for all tasks first, and then to optimize coefficients of the selected features within each task.

Multi-task LASSO can be applied to our problem. Note that, in each region t of brand Wagyu, distinct samples, i.e., heads of beef cattle, are grown up so that we have measured feature sets $\mathbf{X}^{(t)}$ for each region t. As for the trait, we apply the same trait such as BMS, but each region has their own beef cattle, so that the data is expressed as $\mathbf{y}^{(t)}$. By solving multi-task LASSO with the above $\mathbf{X}^{(t)}$ and $\mathbf{y}^{(t)}$, we can obtain the commonly effective feature set among multiple regions within the framework of multi-task LASSO. However, the problem is that multi-task LASSO lacks optimality so that a non-optimal set of features would be selected frequently. In this paper, we try to improve the optimality utilizing correlation coefficients between $\mathbf{y}^{(t)}$ and $\mathbf{x}_f^{(t)}$.

4 Proposed Method

In our method, we select a set of features from F that is commonly effective among multiple measured sets $X^{(t)}$, $t \in T$. Specifically, as LASSO is based on multiple regression analysis, we aim at obtaining a set of features that leads equally high multiple correlation coefficient for every sets $X^{(t)}$.

Our strategy for this is to utilize (not multiple) correlation coefficients between each one feature and the trait variable. There is a tendency that, if the multiple correlation coefficient for a set of features is high, then each of the feature also has a high correlation coefficient. Thus, to select a few features from hundreds or thousands, we first compute the correlation coefficient of each feature, and make preliminary selection. After obtaining several tens of features, we test all the combinations of them and find the best combination of features that has the highest multiple correlation coefficient.

Recall that we denote each region of Wagyu brand by $t \in T$, feature measurement data set by $\mathbf{X}^{(t)}$, and measured trait data set by $\mathbf{y}^{(t)}$. Then, the computational steps of the proposed method is as follows:

Table 1: Notation

Symbol	Description
F	A set of features in input data.
T	A set of region of Wagyu brand.
$\mathbf{x}_f^{(t)}$	Measured feature vector for feature $f \in F$ in region $t \in T$.
$\mathbf{X}_f^{(t)}$	Measured feature matrix that includs all $\mathbf{x}_f^{(t)}$.
$\mathbf{y}^{(t)}$	Measured trait value vector in region $t \in T$.
λ	The regularization parameter in LASSO.
$\beta_f^{(t)}$	The coefficient for feature f and region t.
w_f	L1 norm of $\beta_f^{(t)}$ in terms of region $t \in T$.
$\mathcal{C}_f^{(t)}$	Correlation coefficient between $\mathbf{x}_f^{(t)}$ and $\mathbf{y}_f^{(t)}$.
$\mathcal{C}_f^{(all)}$	Correlation coefficient in feature $f \in F$ with all samples throughout region T.
$\mathcal{J}_f$	Jain's fairness index computed on $f \in F$.
J	A threshold applied to $\mathcal{F}_f$ in step 4.
N	The number of selected features in step 5.
M	The number of features to be selected finally.

1. Computing the correlation coefficient between each feature and the target trait in each region of Wagyu brand.
2. Computing the fairness index of the correlation coefficients among regions, for each feature $f \in F$.
3. Computing the correlation coefficient between each feature and the target trait with all samples of all regions.
4. Removing features from F if the computed fairness index is below the threshold J, and obtain the subset $F' \subset F$.
5. If $|F'|$ is larger than the predefined number N, select top-N features in terms of correlation coefficient computed from all regions, and obtain the subset $F'' \subseteq F'$.
6. Compute the multiple correlation coefficient for every combination of M features in F'', and retrieve the combination that has the highest value.

First of all, notations used in this paper are shown in Table 1.

In step 1, we just compute the correlation coefficients for each feature and region. Specifically, we compute the correlation coefficient of $\mathbf{x}_f^{(t)}$ and $\mathbf{y}^{(t)}$ for each $f \in F$ and $t \in T$, which we denote by $\mathcal{C}_f^{(t)}$.

Step 2 examines the fairness among the computed correlation coefficient of multiple regions. Here we measure the fairness using Jain's fairness index [9], which takes value 1 when all values are the same and takes value n^{-1} in the worst case where n is the number of values. We take the fairness index of the values among regions, so the fairness index is defined as:

$$\mathcal{J}_f = \mathcal{J}(\mathcal{C}_f^{(1)}, \mathcal{C}_f^{(2)}, \ldots, \mathcal{C}_f^{(T)}) = \frac{(\sum_{t \in T} \mathcal{C}_f^{(t)})^2}{n \sum_{t \in T} (\mathcal{C}_f^{(t)})^2}, \quad (3)$$

and this value is computed for each $f \in F$.

Step 3 simply computes the correlation coefficient with all samples of all region with the corresponding trait values. If we define a vector $\mathbf{x}_f^{(all)} = (\mathbf{x}_f^{(1)}, \mathbf{x}_f^{(2)}, \ldots, \mathbf{x}_f^{(T)})$ and $\mathbf{y}^{(all)} = (\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \ldots, \mathbf{y}^{(T)})$, we compute the correlation coefficient of $\mathbf{x}_f^{(all)}$ and $\mathbf{y}^{(all)}$ and denote the value by $\mathcal{C}_f^{(all)}$. We compute $\mathcal{C}_f^{(all)}$ for each $f \in F$.

In Step 4, we apply threshold J on the fairness index $\mathcal{J}_f$ and obtain the subset $F' \subset F$. Thereshold J is preconfigured, and all features f where $\mathcal{J}_f \geq J$ form the set F'. By using the fairness index, we intend to retrieve the features that have fair correlation values for all regions in T.

In Step 5, we just select top-N features in F' in terms of $\mathcal{C}_f^{(all)}$, and obtain the new set F''. This is intended to reduce the computational load of Step 6 within the feasible level represented by N. Note that, to be candidates of the final feature set, the correlation coefficient as well as the fairness should be high. So, we in this paper propose to apply a threshould on fairness first, and then limit the number of features using correlation coefficients.

In Step 6, we finally select the best combination of features from F''. We compute the multiple correlation coefficient for every combination of M features in F'' and output the best combination among them. This would be the optimal combination of M features that have the best multiple correlation coefficient.

5 Evaluation

5.1 Data

We evaluate the proposed method compared with the result of multi-task LASSO. The data consists of 3 regions of branded Wagyu, which we refer region A, B, and C, and each region has 51, 10, and 35 beef cattle (i.e., samples) in the data. Each beef cattle has been grown up in one of the regions, and the 6 economical traits were measured before slaughtered and sold as meat. As mentioned before, the 6 economical traits are CW (Carcass Weight), REA (Rib-Eye Area), RT (Rib Thickness), SFT (Subcutaneous Fat Thickness), YE (Yield Enhancement), and BMS (Beef Merbling Standard). As a result, our data has 6 trait values for each sample from 3 regions.

The feature data set is a proteome expression profile of serum; for each beef cattle, serum is taken with the interval of 3-4 months, which are analyzed by SWATH-MS [10] (Sequential Window Acquisition of all THeoretical fragment ion spectra Mass Spectrometry) method with our own preprocessing treatment. In this method, we got the expression levels of 137 proteins for each sample. As a result, we got 137 protein expression values for 6 periods of time, so we have $137 \times 6 = 822$ features for each sample from 3 regions. After removing the features with null values, we have 580 features to apply the proposed method.

5.2 Method

We applied the proposed method and Multi-task LASSO (MT-LASSO) to the aforementioned data set. Since LASSO is based on the optimization in multiple linear regression, our evaluation criteria is multiple correlation coefficient computed with the selected features. The best-suit method to be compared with the proposed method is MT-LASSO. Because we are intending to retrieve the feature set which is optimal in terms of fairness, we computed the multiple correlation coefficient for each region and examine the variation of them using Jain's fairness index [9]. If the fairness index takes high value, i.e., close to 1, the selected feature set is regarded to be effective in all regions, meaning that the effects of features are not specific to regions, i.e., they are commonly effective elements in Wagyu beef. This meets the objective of our study in this paper.

In the proposed method, we set $M = 3$, $N = 50$, and $J = 0.8$. Namely, we decided to select the best $M = 3$ features to explain each target trait of Wagyu beef. With $M = 3$, the computational time for multiple regression for every pair of 3 features is feasible when $N = 50$ or less (i.e., about 40 minutes per trait with Intel(R) Core(TM) m3-7Y30 CPU). Also, when the threshold value is $J = 0.8$, we obtain as many as 50 features on avarage with each trait. In MT-LASSO, we manually set λ so that we can obtain exactly 3 features for each trait.

We implemented the proposed method with general statistical methods. As for MT-LASSO, although we found MT-LASSO implemented in scikit-learn for Python [12], its implementation does not fit to our case; it only affords the case in which the same feature set is used for multiple objective functions. Thus, we used the LASSO implementation of scikit-learn in place of MT-LASSO. Note that LASSO is equivalent to MT-LASSO in terms of feature selection.

5.3 Results

The results are shown in Table 2, in which the performance of the proposed method and MT-LASSO are compared for 6 economical traits. Note that the results of MT-LASSO shown in the Table is not the value obtained by MT-LASSO. In order to fairly compare the results with the proposed method, we compute the multiple regression coefficients with the features selected by MT-LASSO, and use the values in the Table.

First of all, we focus on that the multiple correlation coefficients with all samples are better with the proposed method than with MT-LASSO in all of 6 traits. This means that the proposed method is better than MT-LASSO in terms of optimization performance. This also means that the preliminary selection of $N = 50$ features based on single correlation coefficient works effectively to capture the commonly effective features within 50 features.

Second, we see that the fairness index of multiple correlation coefficients among regions are better with the proposed method than with MT-LASSO in 5 out of 6 economical traits. Namely, the multiple correlation coefficients in different region take relatively closer values in the proposed method than MT-LASSO. This means that the retrieved features of the proposed method effects more commonly on different region than MT-LASSO.

From above, we conclude that the proposed method performs better than MT-LASSO not only in optimality of the

Table 2: Evaluation Results

Trait	Method	Multiple Correlation Coefficient				Fairness Index
		All	Region A	Region B	Region C	
CW (Carcass Weight)	Proposed	0.4489	0.5242	0.4956	0.5370	0.9989
	MT-LASSO	0.3692	0.4104	0.6140	0.5706	0.9736
REA (Rib-Eye Area)	Proposed	0.4596	0.2607	0.6739	0.5785	0.8907
	MT-LASSO	0.4233	0.4567	0.7330	0.3847	0.9243
RT (Rib Thickness)	Proposed	0.5558	0.5835	0.8145	0.5138	0.9609
	MT-LASSO	0.4913	0.5191	0.7978	0.4389	0.9354
SFT (Subcutaneous Fat Thickness)	Proposed	0.4691	0.4081	0.3861	0.6147	0.9541
	MT-LASSO	0.4150	0.3102	0.3884	0.5843	0.9322
YE (Yield Enhancement)	Proposed	0.4487	0.3793	0.5202	0.4802	0.9837
	MT-LASSO	0.3639	0.4309	0.6276	0.3082	0.9230
BMS (Beef Marbling Standard)	Proposed	0.4951	0.5350	0.7020	0.3990	0.9509
	MT-LASSO	0.4847	0.6104	0.7699	0.2619	0.8694

retrieved features but also in terms of commonness of effects among different regions.

6 Conclusion

In this paper, we proposed a method to find the optimal feature sets from very large number of features, which are commonly related to the target trait in each of the multiple groups of data, within feasible computational time. Although multi-task LASSO has already been proposed to tackle this problem, it has a problem that it does not output optimal feature sets frequently, and also that it does not care much about the fairness of the effect in each group.

To solve the problem, we proposed a method to previously select candidate features, and then examine all the combinations of the features to retrieve the optimal and fair-effect feature set. Our previous selection method computes the correlation coefficient between feature and trait values for each feature and each group. Then, we compute the average of them with all groups as well as the fairness of them among groups. By selecting features that have both high average and fairness, we obtain candidate features that commonly effects on the target trait in every groups.

We apply the proposed method to the data set of Wagyu proteome profile data set that includes samples from multiple regions of branded Wagyu. As a result, we found that the proposed method outperforms multi-task lasso in both optimality, and fairness of effects among multiple regions. From the viewpoint of Wagyu breeding, our trial is to find the protein set that are truly related to beef quality by excluding the regional effects such as breeding methods of cattle specific to the region. With the proposed method, we would be possible to select the proteins related to 6 economical traits more efficiently than the conventional multi-task LASSO scheme.

As future work, we first collect more evidence on the excellence of the proposed method through additional evaluation. We also would try the case of larger M, in which feasible N is smaller so that obtaining optimal solution might require more sophisticated techniques.

ACKNOWLEDGMENT

This work was partly supported by “the Program for Promotion of Stockbreeding”of JRA (Japan Racing Association).

REFERENCES

[1] R. Tibshirani, Regression Shrinkage and Selection via the Lasso, Journal of the Royal Statistical Society, Series B, 58(1), pp.267–288 (1996).

[2] G. Obozinski, B. Taskar, and M. Jordan., Multi-task feature selection, In Technical Report, Department of Statistics, University of California, Berkeley, 2006.

[3] S. Hara and T. Maehara, “Enumerate Lasso Solutions for Feature Selection,” In Proc. AAAI2017, 2017.

[4] Japan Meat Grading Association, “The Manual of Standard in Dealing Pork and Beef,” 2001 (In Japanese).

[5] A Guide of Japanese Black Cattle Sires, http://liaj.lin.gr.jp/index.php/detail/data/m/803237096 (referred in May 2017) (In Japanese).

[6] Wagyu Registry Association,“Compliation of Sires of Japanese Black Cattle,” 2003 (In Japanese).

[7] N. D. Cameron, “Selection Indices and Prediction of Genetic Merit in Animal Breeding,” CAB International (1997).

[8] A. Oka, T. Dohgo, M. Juen, and T. Saito, “Effects of vitamin A on beef quality, weight gain, and serum concentration of thyroid hormones, insulin-like growth factor-I, and insulin in Japanese black steers,” Animal Science and Technology (1998).

[9] R. Jain, D.M. Chiu, W. Hawe, “A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems,” DEC Research Report TR-301, 1984.

[10] C. Ludwig, L. Gillet, G. Rosenberger, S. Amon, B.C. Collins, R. Aebersold, “Data- independent acquisition- based SWATH- MS for quantitative proteomics: a tutorial,” Molecular Systems Biology (2018) 14, e8126, DOI 10.15252/msb.20178126, 2018.

[11] R-Project, https://www.r-project.org/

[12] Scikit-learn, https://scikit-learn.org/

Keynote Speech 3:
Dr. Katsuhiko Kawazoe
(Senior Vice President / Head of Research and Development Planning, Nippon Telegraph and Telephone Corporation)

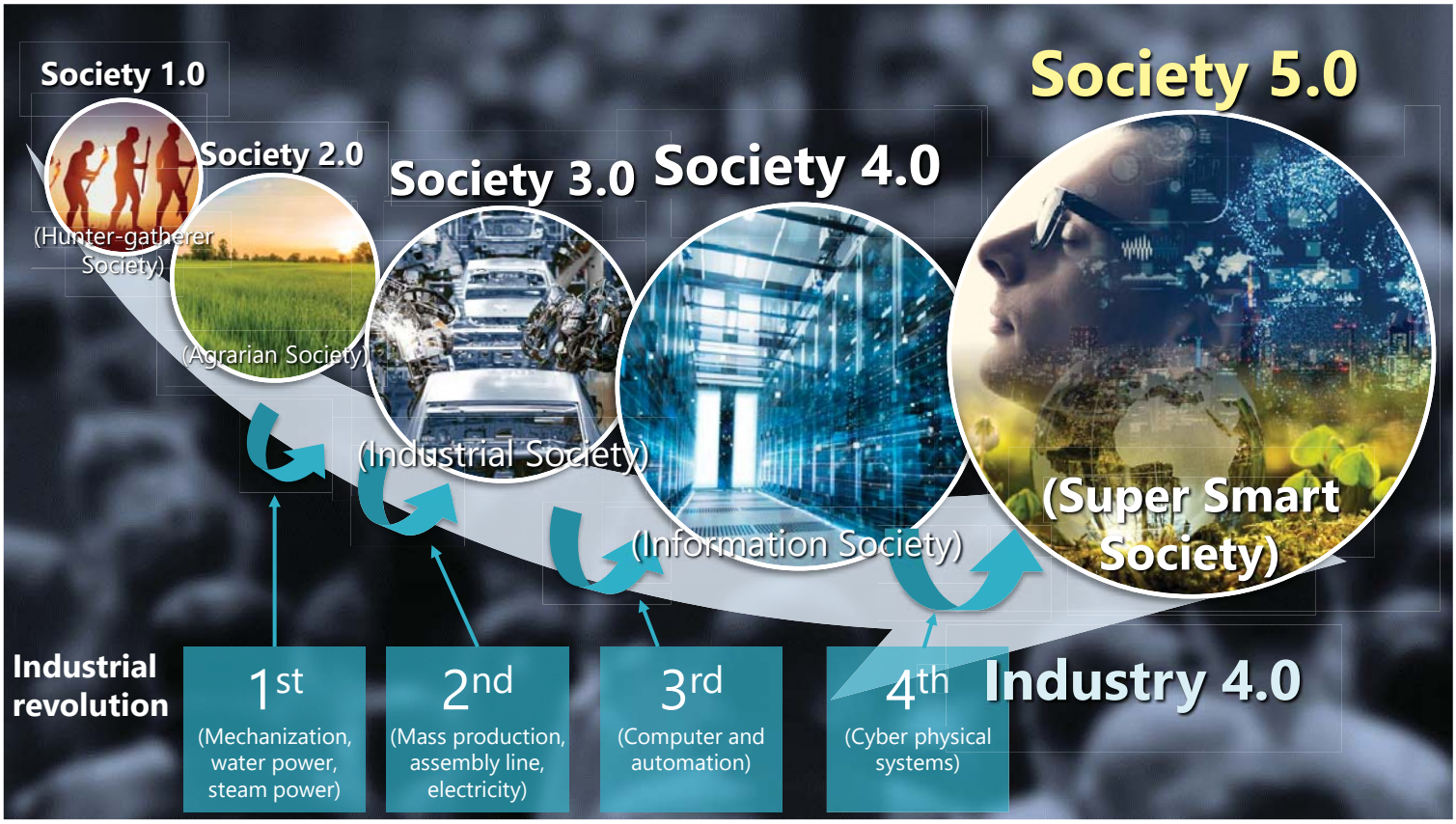
NTT

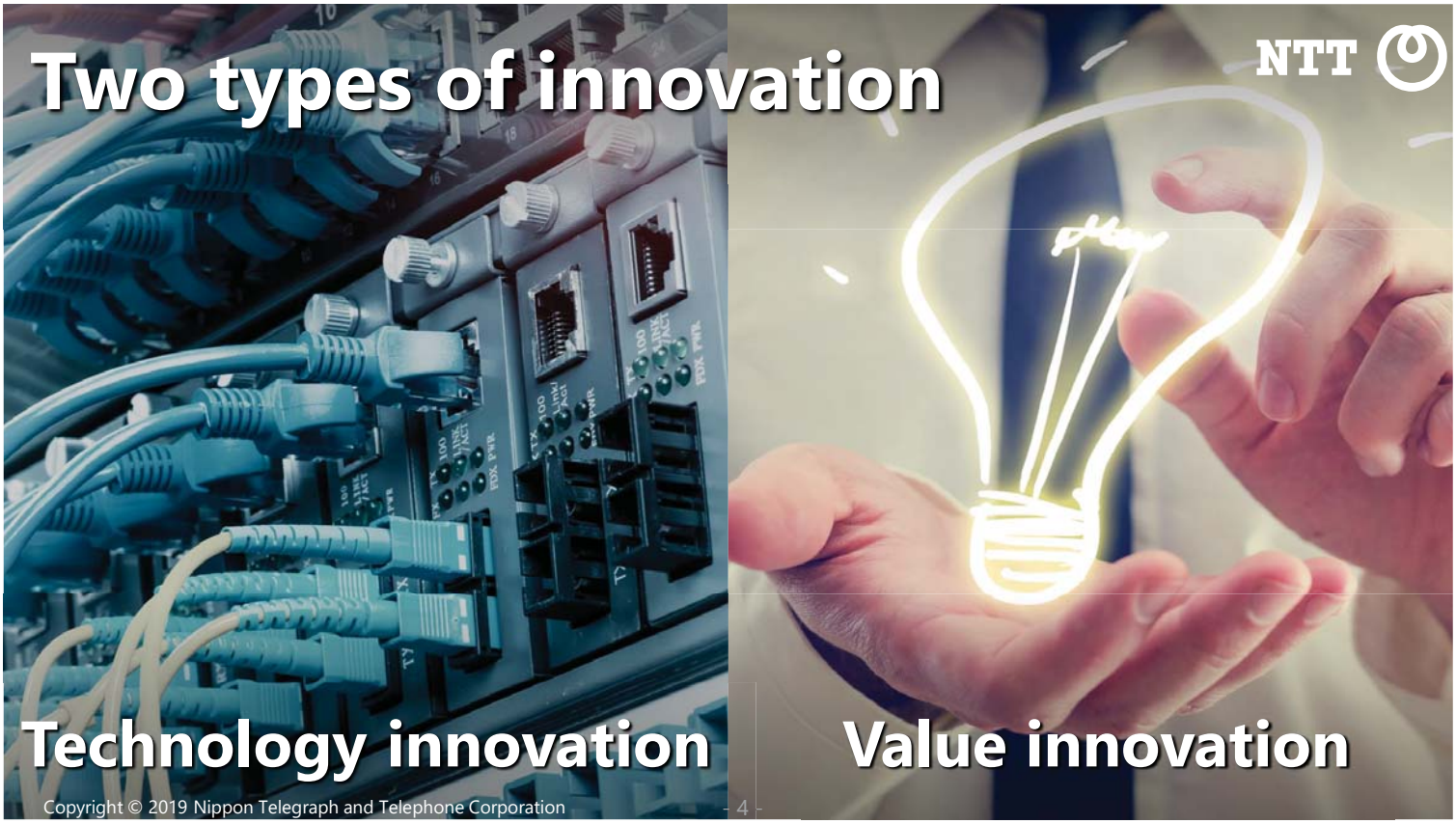
Digital to Natural

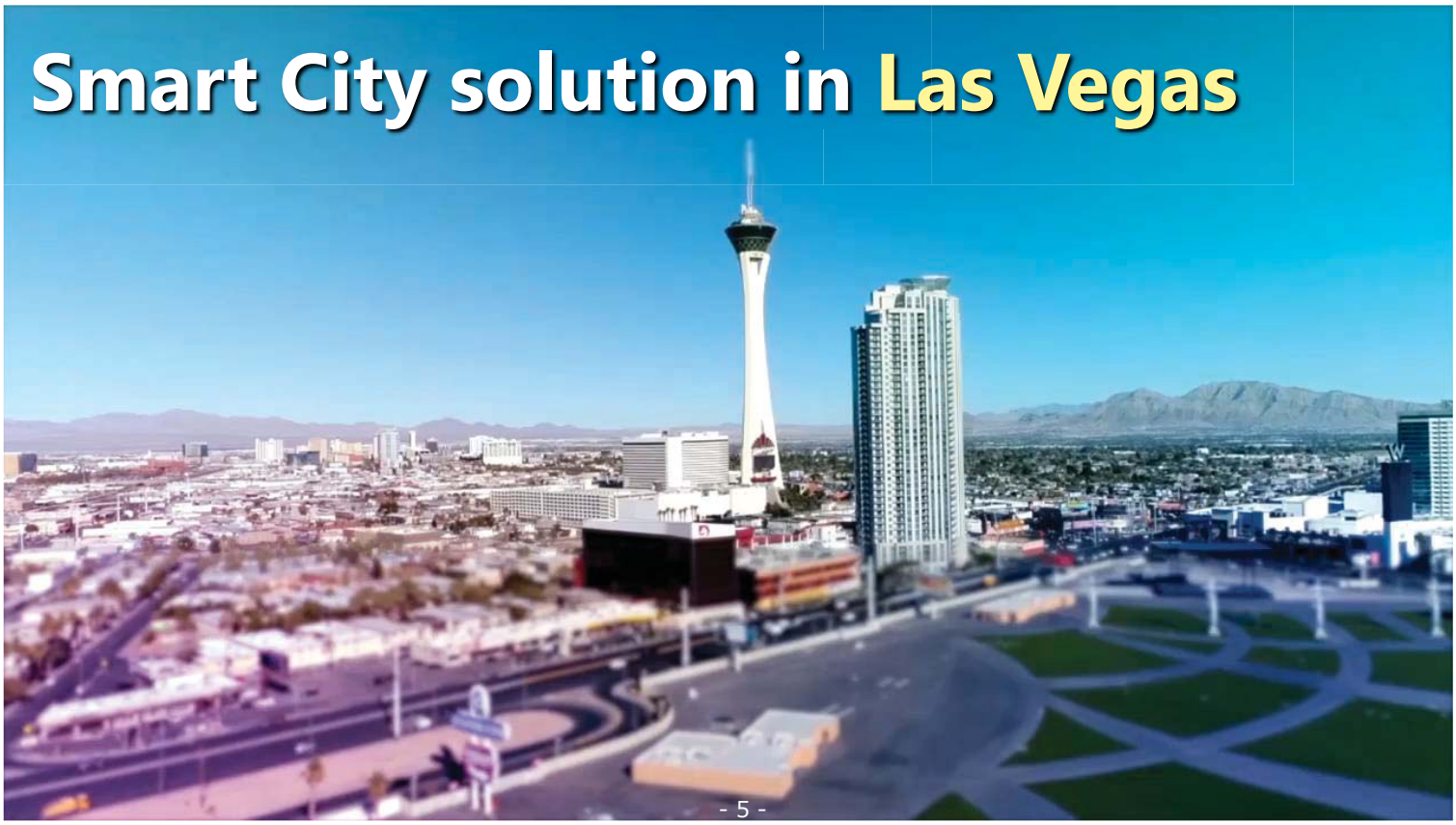
- Innovation for Smart World -

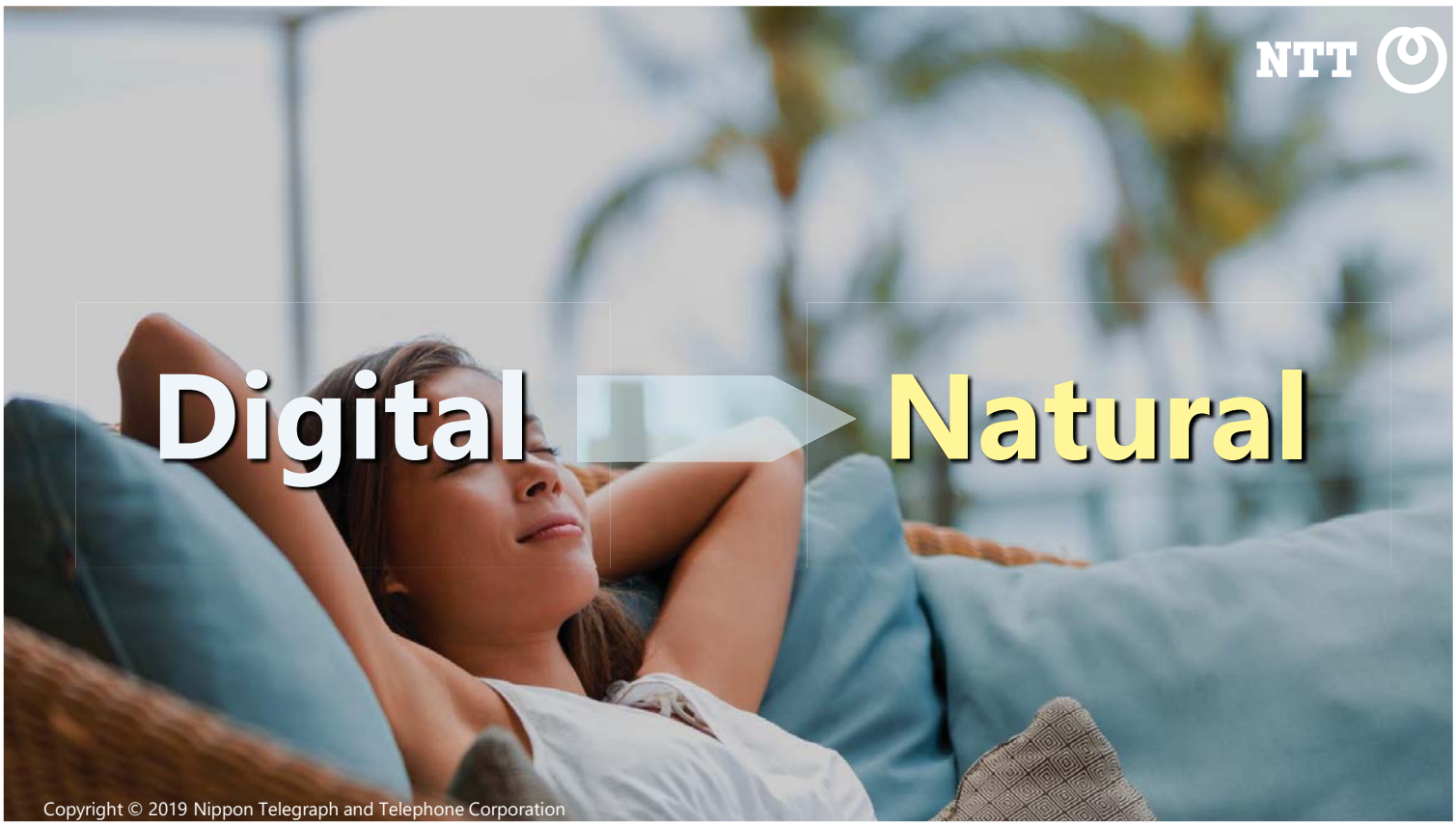
NIPPON TELEGRAPH AND TELEPHONE CORPORATION
Senior Vice President,
Head of Research and Development Planning

Katsuhiko Kawazoe






Smart City solution in Las Vegas
- 5 -


NTT
Digital
Natural
Copyright © 2019 Nippon Telegraph and Telephone Corporation


generous
comfortable
NTT
intuitive
considerate
human friendly
subconscious
heart touching
Natural
subliminal
emotional
flexible
ubiquitous
autonomous
adaptable
Copyright © 2019 Nippon Telegraph and Telephone Corporation
- 7 -


Natural discussion with robots
- 8 -

Crossmodal voice / face conversion

NTT

Man's voice

Generating a face
to match the input voice

Man's voice

Woman's voice

Converting a voice
to match the input face

- 9 -

Heart-warming elderly care

Supporting the daily lives of elderly people with devices tailored to help those with dementia

"Dementia gear"

What was today's breakfast?

What was today's breakfast?

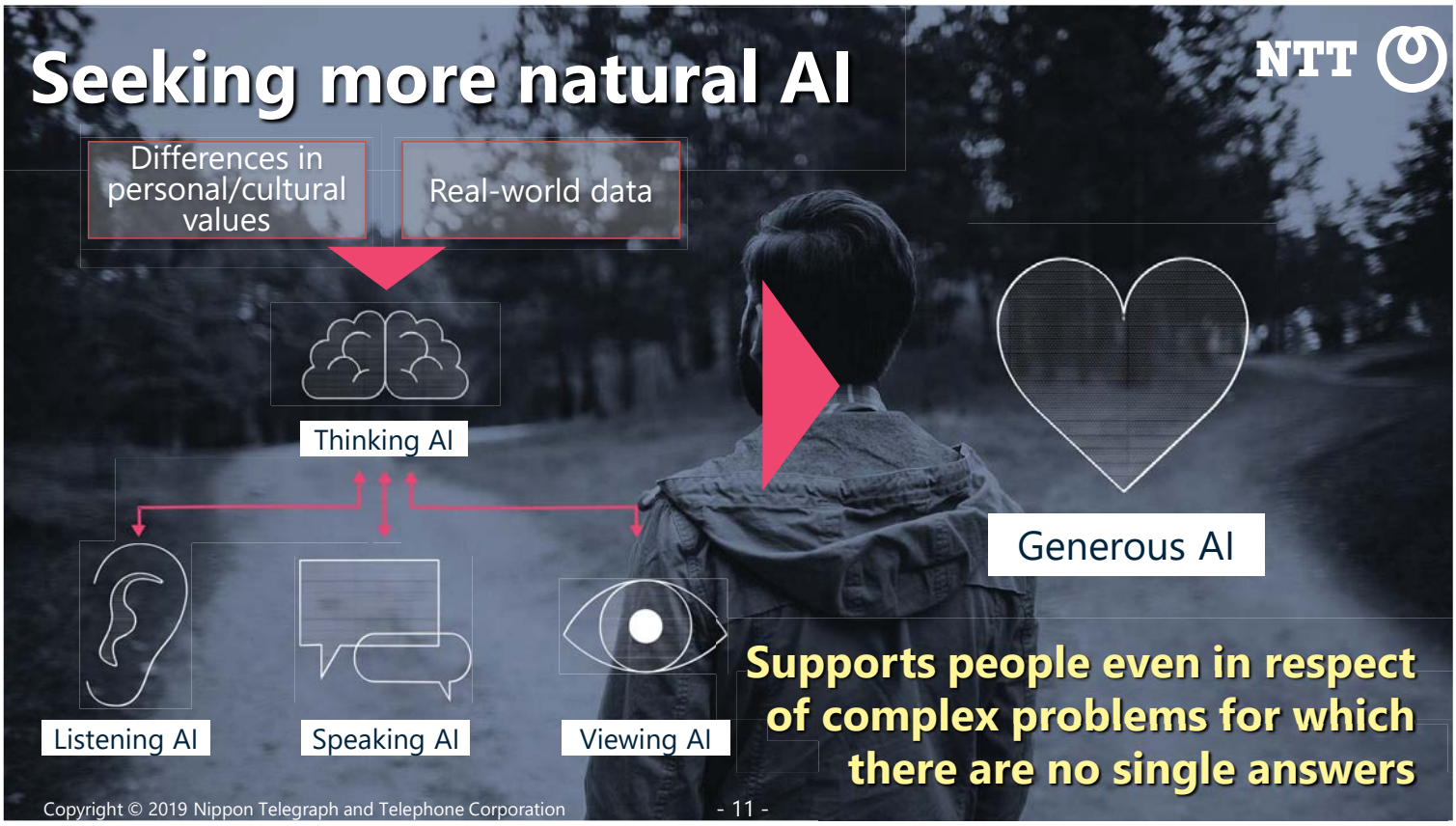
I had toast.

I'll support his conversation in an amiable way

Personal twin

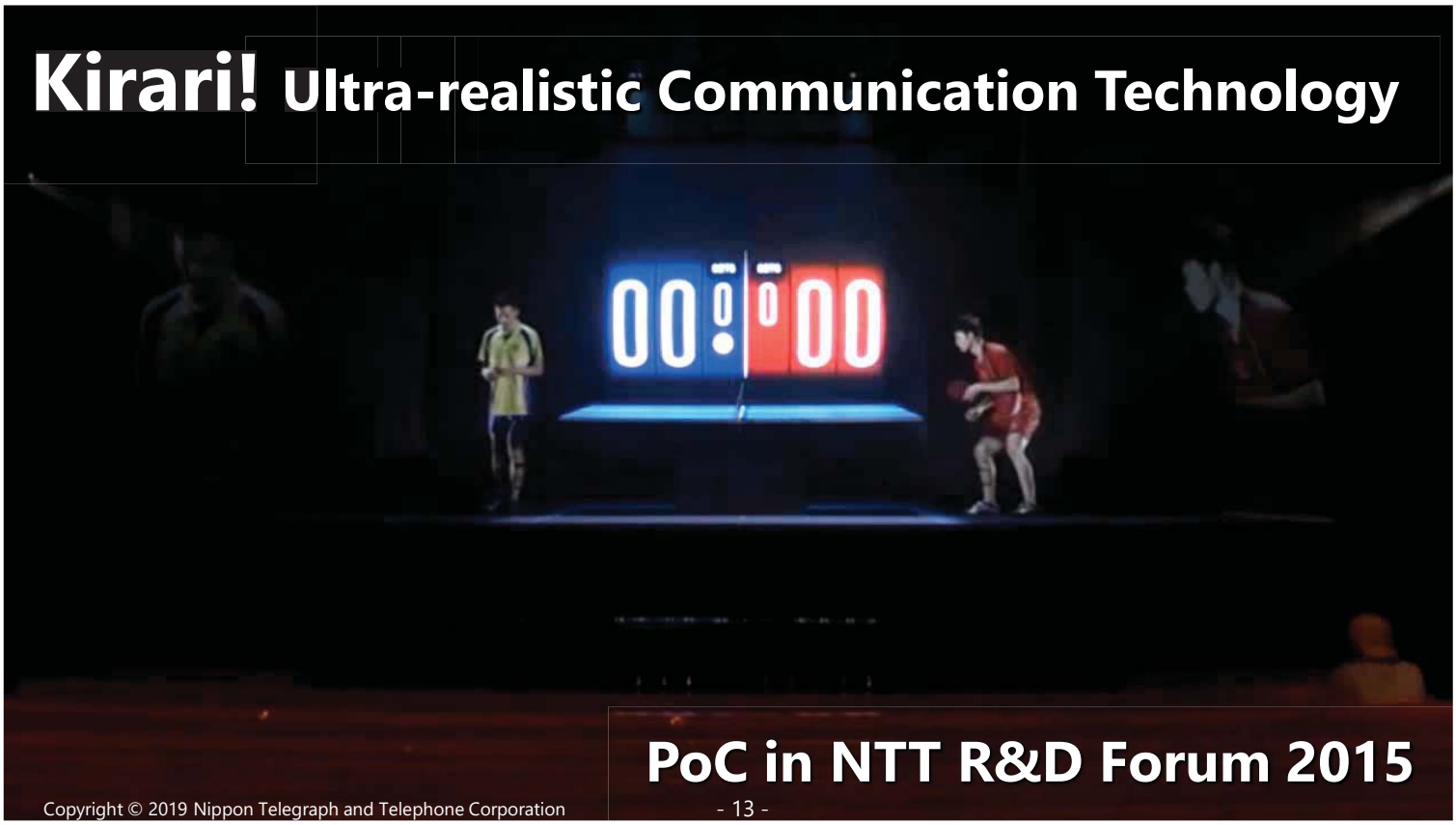
- replaying phrases that are difficult to hear
- using his own voice to answer difficult questions on his behalf
- giving encouragement to answer easy questions

- 10 -

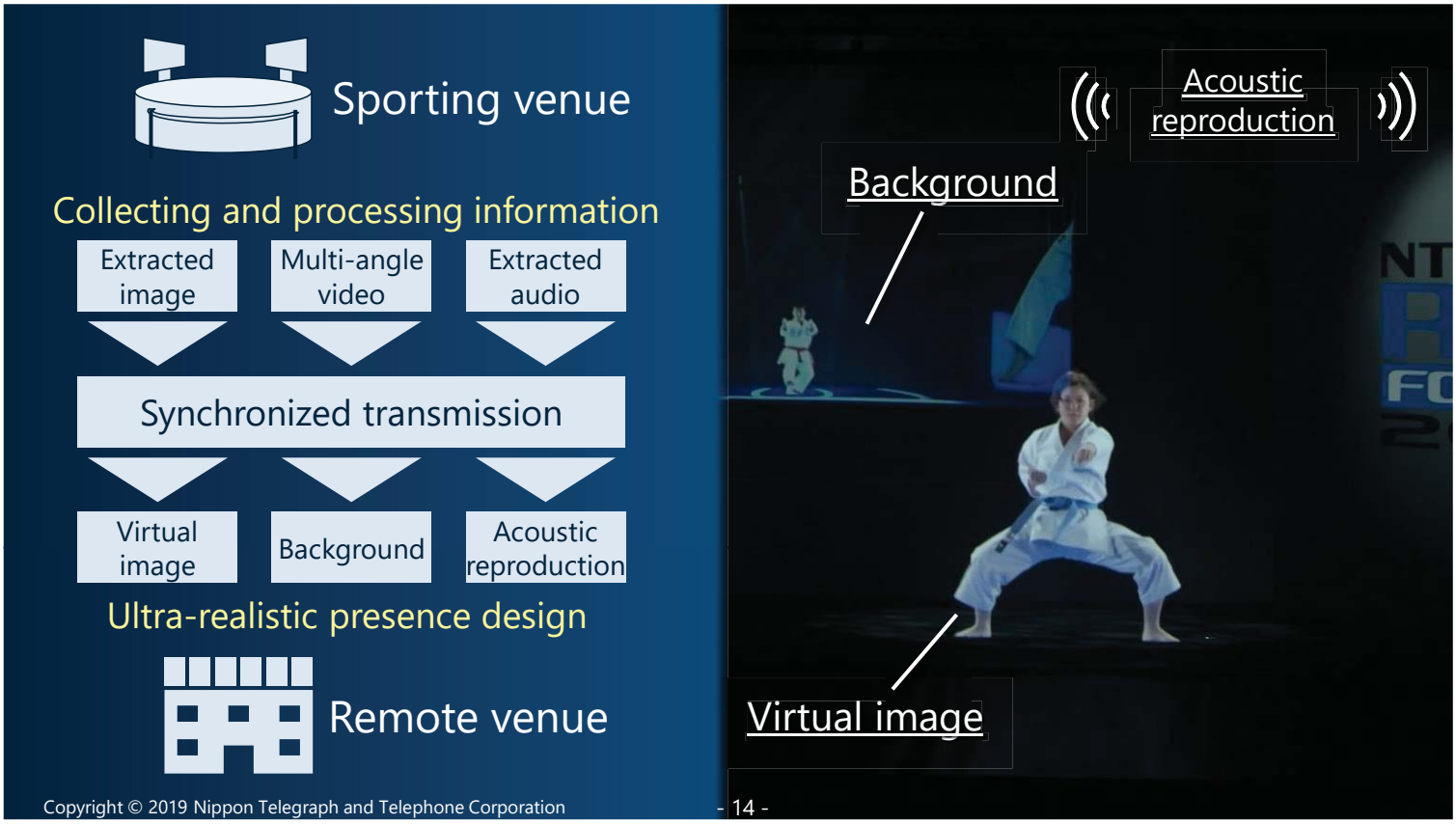

Seeking more natural AI
NTT
Differences in personal/cultural values
Real-world data
Thinking AI
Generous AI
Listening AI
Speaking AI
Viewing AI
Supports people even in respect of complex problems for which there are no single answers
Copyright © 2019 Nippon Telegraph and Telephone Corporation
- 11 -


Natural excitement by ultra-realistic communication
NTT
You can experience the sensation of being at the event site, no matter where you are.
Copyright © 2019 Nippon Telegraph and Telephone Corporation
- 12 -

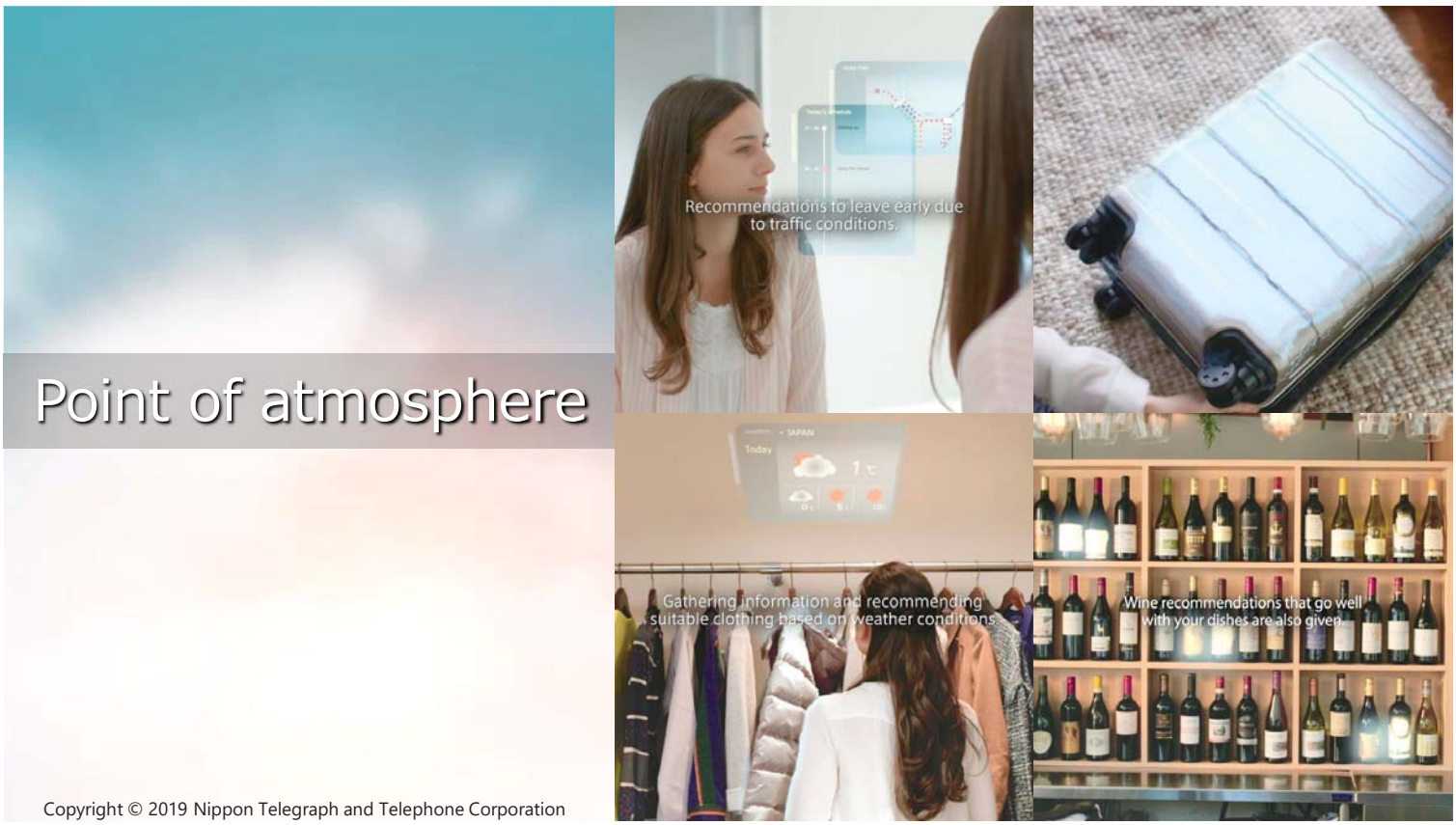
Kirari! Ultra-realistic Communication Technology



PoC in NTT R&D Forum 2015




NTT
1950
1970
2000
2010
09:30
?
What's next ?
Copyright © 2019 Nippon Telegraph and Telephone Corporation


Point of atmosphere
Recommendations to leave early due to traffic conditions.
Gathering information and recommending suitable clothing based on weather conditions.
Wine recommendations that go well with your dishes are also given.
Copyright © 2019 Nippon Telegraph and Telephone Corporation

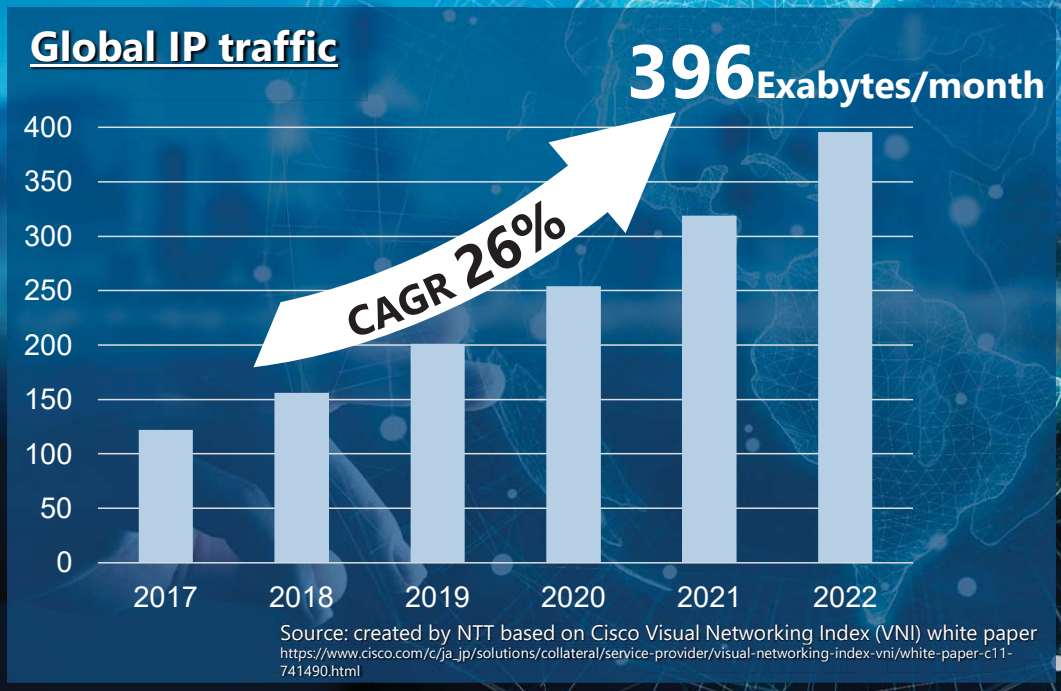
The future network

NTT

blending seamlessly into society

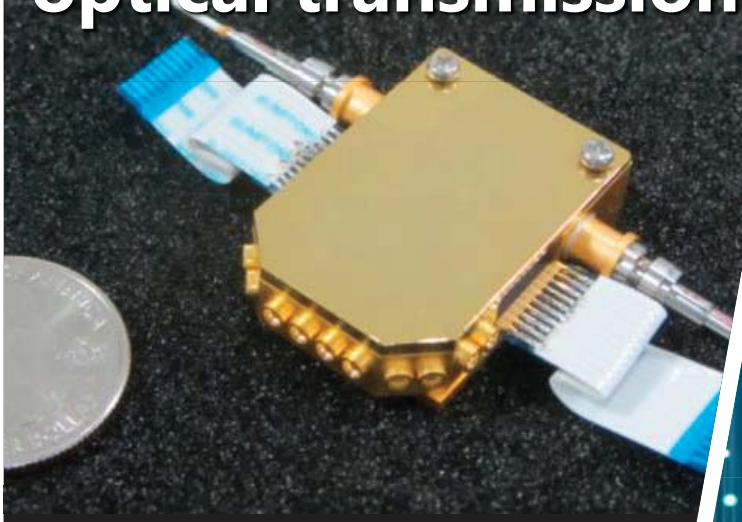
Increase in traffic

NTT

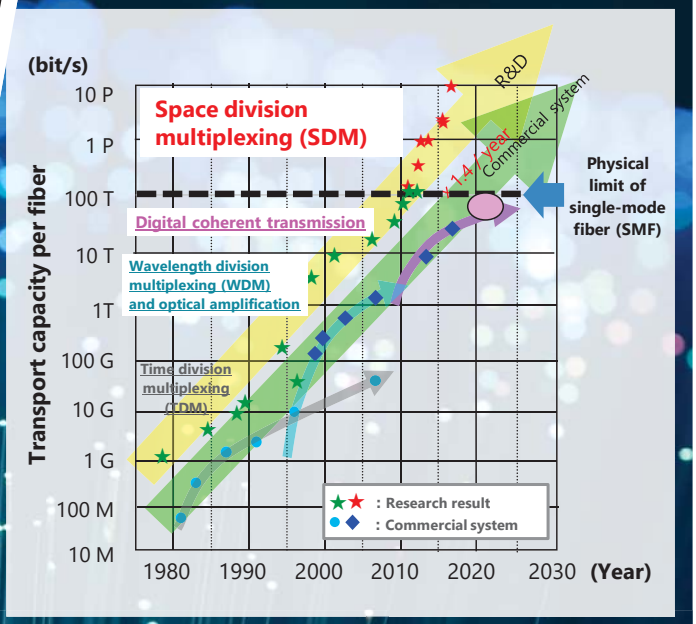


Source: created by NTT based on Cisco Visual Networking Index (VNI) white paper
https://www.cisco.com/c/ja_jp/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html

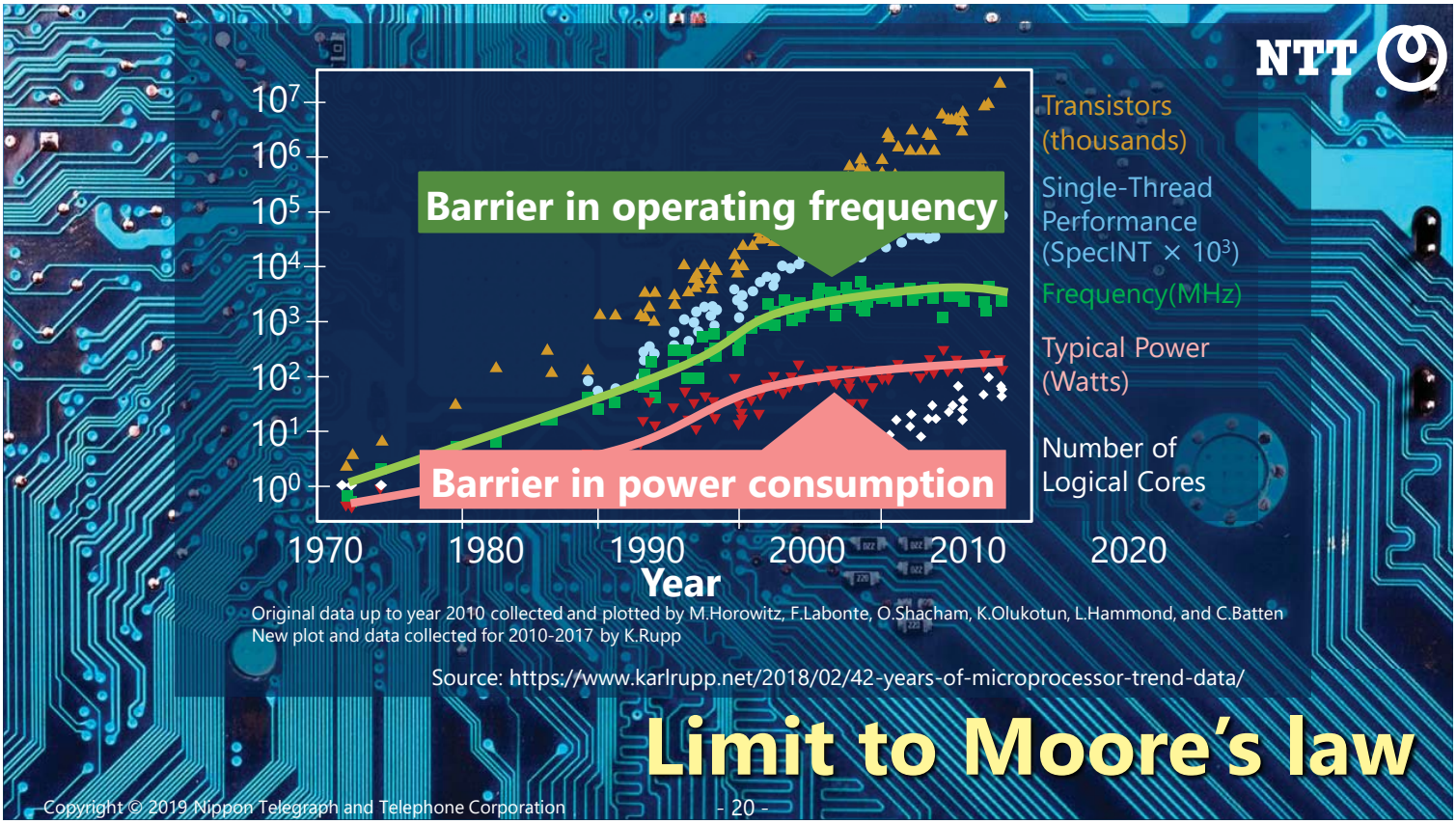
1Tbps / λ WDM optical transmission



800km, unparalleled long distance transmission



Evolution of high-capacity optical transport network



Limit to Moore's law

NTT

IOWN

Innovative Optical and Wireless Network

- All Photonics Network
- Digital Twin Computing
- Cognitive Foundation

All-photonics network

NTT

✓ is a game changer ushering in a new information infrastructure

Where	From network to terminals, including optical fibers, transmission systems and semiconductors
Technology	Introduce photonics-based technologies everywhere
Goal	**All-photonics network** A game changer transforming the electronics world into the photonics world

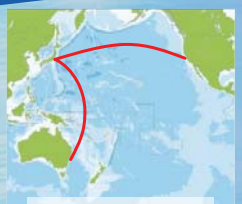
From electronics to photonics

NTT

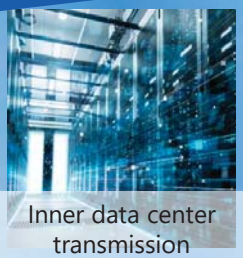
Transmission using photonics
Processing using electronics

Processing using optoelectronic integration

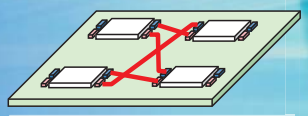
Shortening distance of optical transmission



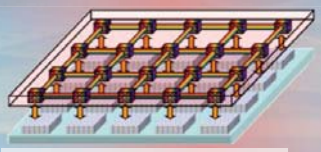
Long distance transmission (>1 km)



Inner data center transmission (< several hundred meters)

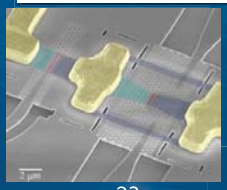


Inter-chip transmission (mm - cm)



Inter-core transmission
Inner-chip processing
(< 1mm)

Press release on April 16, 2019



Ultra-low-energy and high-speed optical transistor
Nature Photonics, Vol. 13(7).

Advantages of all-photonics network

NTT

Lower power consumption

Power efficiency:
100 times higher*

Transmission media
Optical fiber cables

Transmission systems
Light (wavelengths) throughout

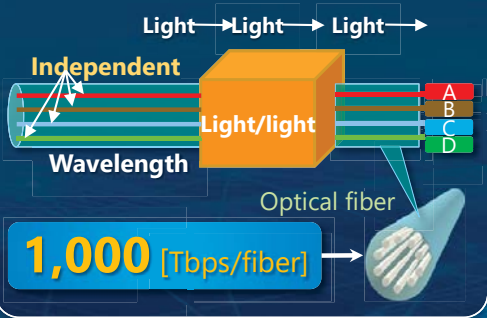
Information processing platform
Optoelectronic integrated devices

*Target power efficiency for the photonics part

High quality and high capacity

Transmission capacity:
125 times higher*

- Wavelengths (optical signal)



1,000 [Tbps/fiber]

*Target communication capacity per fiber

Low delay

End-to-end delay:
1/200*

- Transmission per wavelength
- No queueing
- No data compression

Wavelength A: **High-capacity video (uncompressed)**

No processing delay

Wavelength B: **Speech**

*Target delay for uncompressed video traffic within the same prefecture

Digital twin computing

NTT

■ Before

Digital twins used individually

Autono -mous driving

Robotics

Medical treatment

Digital twins

Things

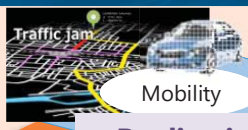
Humans

■ After

Exchange and expansion of abilities

Multifaceted personal decision making

Prediction and containment of disease spread



Mobility

City

Replication, fusion and exchange

Digital twin computing

Digital twins

Things

Humans

Digital twin with IOWN

NTT

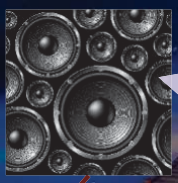
Conventional Digital twin

"Natural harmonic"

Ultra-realistic sports viewing with IOWN

NTT

Uncompressed, high-definition video transmission



Real-time reproduction of various sounds in a stadium

Wavelength A : High-capacity video

Wavelength B : Multipoint voice

Wavelength C : Five-sense data

Wavelength D : Information processing

Synchronous reproduction of the sense of touch in addition to seeing and hearing

Simplification of devices (processing at clouds)

NTT

IOWN

(Innovative Optical and Wireless Network)

IOWN Innovation Forum (provisional)

NTT seeks to make the IOWN concept a reality through collaboration with experts and global partners from a wide range of research and technology fields.

Session 6:
Systems and Services
(Chair: Tomoki Yoshihisa)

A Review Assistance System for Class Diagram with Voice Assistance based on NLP

Masashi Nakamura†, Kozo Okano†, Shinpei Ogata†, and Toshifusa Sekizawa‡

†Faculty of Engineering, Shinshu University, Japan
‡School of Engineering, Nihon University, Japan
{okano, ogata}@cs.shinshu-u.ac.jp
sekizawa@cs.ce.nihon-u.ac.jp

Abstract - Software development needs variety of design documents including requirements, use-case diagrams, class diagrams, communication diagrams, and state-machine diagrams. These documents are reviewed by software engineers before start coding. The review step is an important step to reduce system bugs. It, however, requires human skills and is a time-consuming task. Several approaches including formal approaches are said to be promising techniques to reduce the cost. However, many of the real development engineers rely on a conventional review process. This paper proposes a review assistance system for class diagrams. It analyzes a similarity of each pair of classes of the target diagrams using words used as attributes in the target classes. Based on the analysis of morpheme analysis, it calculates a similarity of the classes. If two classes have a high value of the similarity, we can conclude that the classes might have the common parent class. The proposed system automatically identifies the common class and produces synthesized voice that notifies the engineers. The voice assistance is said to have power superior to text notification for engineers to help finding room for improvement. We implemented a proto-typed system of our proposed method and have checked its usability through some small but real examples of class diagrams. The results show the advantage of our proposed method.

Keywords: Class Diagram, NLP, Review, Refactoring

1 INTRODUCTION

In development of large-scale systems where multiple engineers participate, Unified Modelling Language (UML) has the advantage as the common language for models [1]. UML defines variety of diagrams. Each of them has its own purpose. Some of important diagrams are use-case diagrams, class diagrams, communication diagrams, and state-machine diagrams.

Usually these diagrams are reviewed by software engineers at the system analysis or system design phase. The review step is an important step to reduce system bungs. In general, review process is performed by engineers based on face-to-face base manner. Some documents focused are viewed by the engineers. An engineer resposible to the documents explains the contents of the documents by oer, performed manually and step by step manner.

This paper, aiming at practical use as a review assistance, proposes a method realizing as a plug-in of astah * professional [6]. Astah * professional is software for developing UML diagrams. In addition, in order to confirm whether it is possible to conduct an effective review by the proposed method, the system is applied to multiple actual class diagrams to perform evaluation.

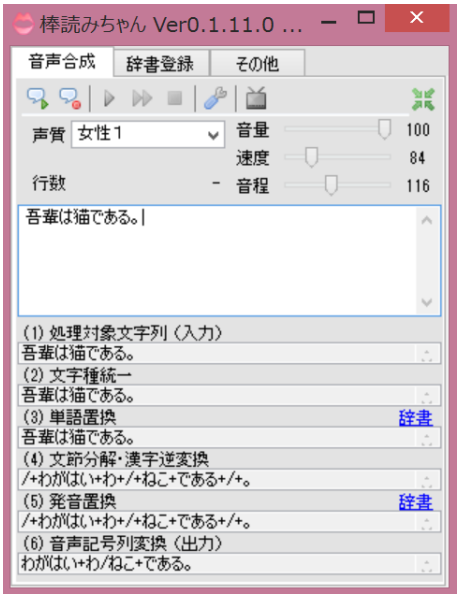


Figure 1: Execution of BouyomiChan

This paper is organized as follows. Section 2 explains terms related to the technology used for development. Section 3 gives an overview of the developed system. Section 4 describes experiments for evaluating the system, the results, and discussion. Section 5 mentions related work. Finally, Section 6 states the conclusion and the future work.

2 PRELIMINARIES

2.1 Speech Synthesis

Speech synthesis is a technique to create automatically human voices from the text. Speech synthesis is used in a wide range of fields such as text-to-speech. One of concrete examples is a singing voice synthesis system[7], VOCALOID [8]. As another example, WikiTalk [9] reads Wikipedia articles. In this report, BouyomiChan [10] was used as speech synthesis software. Figure 1 shows an example of how BouyomiChan works. BouyomiChan is a tool to read Japanese sentences including kanji using speech synthesis technique. It has excellent extensibility. It can be linked with external programs via socket communication.

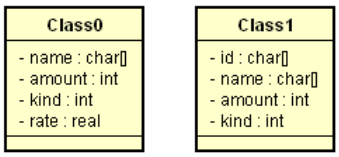


Figure 2: Class Diagram 1

2.2 Morphological Analysis

Morphological analysis [11] is a method to divide sentences in natural language into morphemes and discriminate parts of speech.

English and other European languages use a space character to separate words. In these languages, a word is almost the same as a morpheme. Therefore, it is easy to analyze morphemes in European languages. However, Japanese does not use the space character as a separator. Thus, morphological analysis is not easy task for Japanese.

In order to calculate similarity using Jaccard coefficients, it is necessary to perform morphological analysis of the target character string. If the original string to be analyzed is "wagahaiwanekodearu (I am a cat)," it is divided into morphemes as "wagahai/wa/neko/de/aru." In this report, we used Kuromoji [12] as a Japanese morphological analysis tool. Kuromoji is an open source morphological analysis engine developed in Java, which is particularly easy to handle in Java development.

2.3 Jaccard Coefficient

Jaccard coefficient is one of the coefficients [13] used to calculate the similarity of documents and character strings, in the field of natural language processing. It represents the proportion of common elements in the two sets.

The formula for calculating the Jaccard coefficient is shown in Equation 1. When the sets of A and B consist of "apple, orange, and grape" and "apple, orange, and strawberry." Then $\mid A \cap B \mid = 2$, and $\mid A \cup B \mid = 4$, thus Jaccard coefficient is 0.5

$$J(A,B) = \frac{\mid A \cap B \mid}{\mid A \cup B \mid} \tag{1}$$

The value ranges from 0 to 1. If the value is nearly equal to 1, we think the sets are similar. As an element of the set, we usually use a word or a morpheme of the target sentence or the phrase to analysis.

2.4 Class Diagram and Refactoring

A Class Diagram is one of important diagrams defined in UML.

A class diagram depicts a class as a box with its name, its attributes, and its operations. Figure 2 shows an example of a class diagram. Figure 2 shows two classes, namely class0 and class1. They have no operations for the ease of explanation.

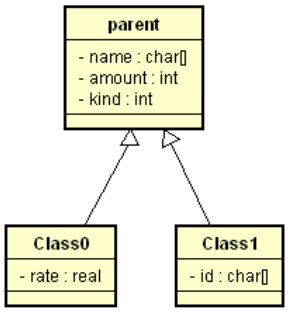


Figure 3: Revised Class Diagram 1

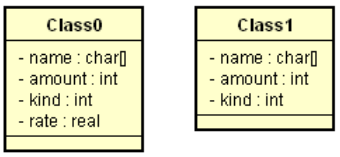


Figure 4: Class Diagram 2

They have many common attributes. In such a case, usually engineers revise the class diagram as shown in Figure 3. This kind of revision is one of important operation of refactoring [14].

In Figure 3, we create a new class namely parent. The revised Class0 and Class1 have only their own attributes and the all of common attributes are moved into the parent class. To indicate that class0 and class1 inherit the common attributes from parent, the revised class diagram has a special relation between parent class and class0 (or class1). The relation is called "generalization relation." We call this operation "creation of generalization relation."

Let's see another case shown in Figure 4.

Here, all attributes of class1 are included in class0. In such a case we can perform yet another refactoring operation as shown in Figure 5. We call this type of operations "creation of inclusion relation."

From the definition, if inclusion relations can be created, then also generalization relations can be created.

In such a case, usually engineers choose the suitable relations.

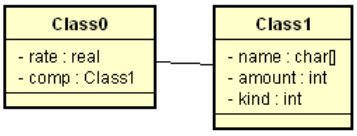


Figure 5: Revised Class Diagram 2

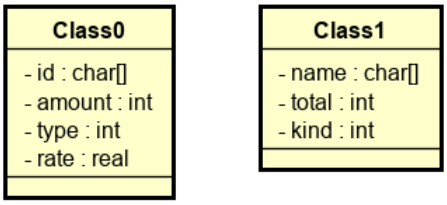


Figure 6: Class Diagram 3

3 OVERVIEW OF THE PROPOSED REVIEW ASSISTANCE SYSTEM

3.1 Problem Statement

Let us consider the class diagram shown in Figure 6.

Terms “id,” “type,” and “total” might be “name,” “kind,” and “amount”, respectively due to lack of confirmation of the engineers. If the terms “id” and “name”, “kind” and “type”, and “amount” and “total”, are the same meaning respectively, then we can create generalization relation.

Of course, however, these terms might be chosen as different words on purpose.

Therefore, engineers have to review carefully for each pair of classes to check the possibility of creation of generalization relation or inclusion relation by considering the possibility of amigos selection of terms.

Such amigos occurs also class diagrams written in Japanese. In the early stage of design, conceptual models are usually created by engineers for the purpose of understanding the target system. In the conceptual models, class diagram usually written in the native language of engineers, such as Japanese.

From the above consideration, we set our goal to assist Japanese engineers to find candidates of generalization relations or inclusion relations in the conceptual class diagram to review.

3.2 The Proposed Method

The proposed method is shown below.

1. It extracts class names, class operations and class attributes (We call both of class operations and class attributes, elements below) from a class diagram.

2. It disassembles each element into morphemes by Kuromoji.

3. It calculates the similarity between each pair of elements using Jaccard coefficient.

4. It filters the similarity based on some threshold.

5. It shows the results by text and also speech voice by speech-synthesis.

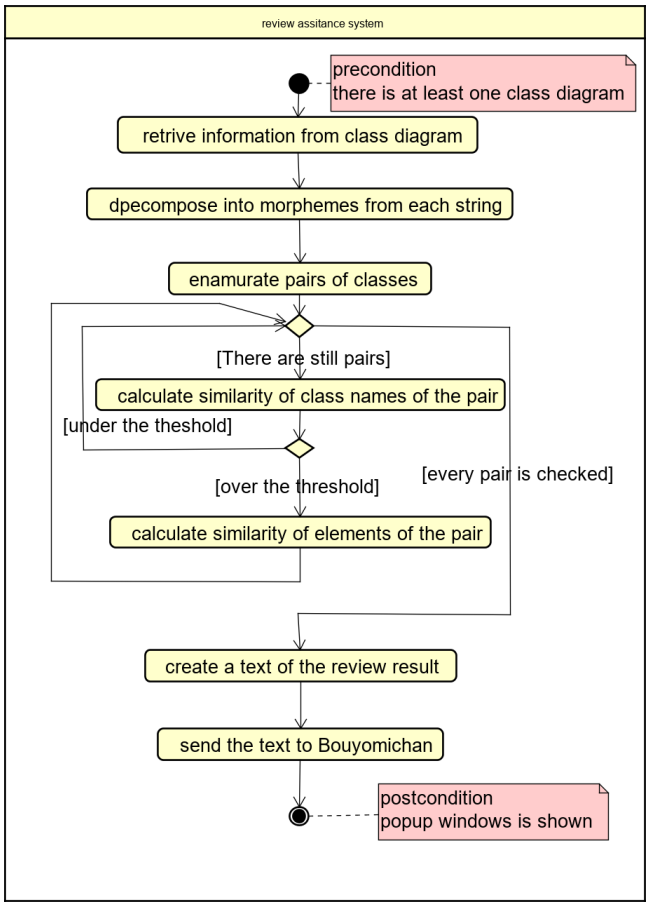


Figure 7: Flow of Review Assistance System

Table 1: Environment

OS	Windows 8.1 64bit
Language	Java 8.1
framework	Astah* Plug-in SDK

3.3 Implementation

Table 1 shows the implementation environment.

We have implemented the proposed method as a plug-in for astah * professional. Figure 7 shows an activity diagram describing the operation of the review assistance system. Figure 7 does not include the behavior of BouyomiChan.

Whenever you draw a class diagram, the plug-in is executed at an arbitrary timing. The result of the review is shown both of voice and a pop-up window. Figure 8 shows an example of a review.

BouyomiChan starts at the same time as the plug-in is executed.

We used astah * API [15] to extract class names, attributes, and operations from class diagrams. The astah * API is a set of Java interfaces for developing application software that utilizes astah * model data.

The threshold for similarity filtering is set statically in advance. In a class diagram with a large number of classes, it is considered that a lot of calculations of similarities of elements need.

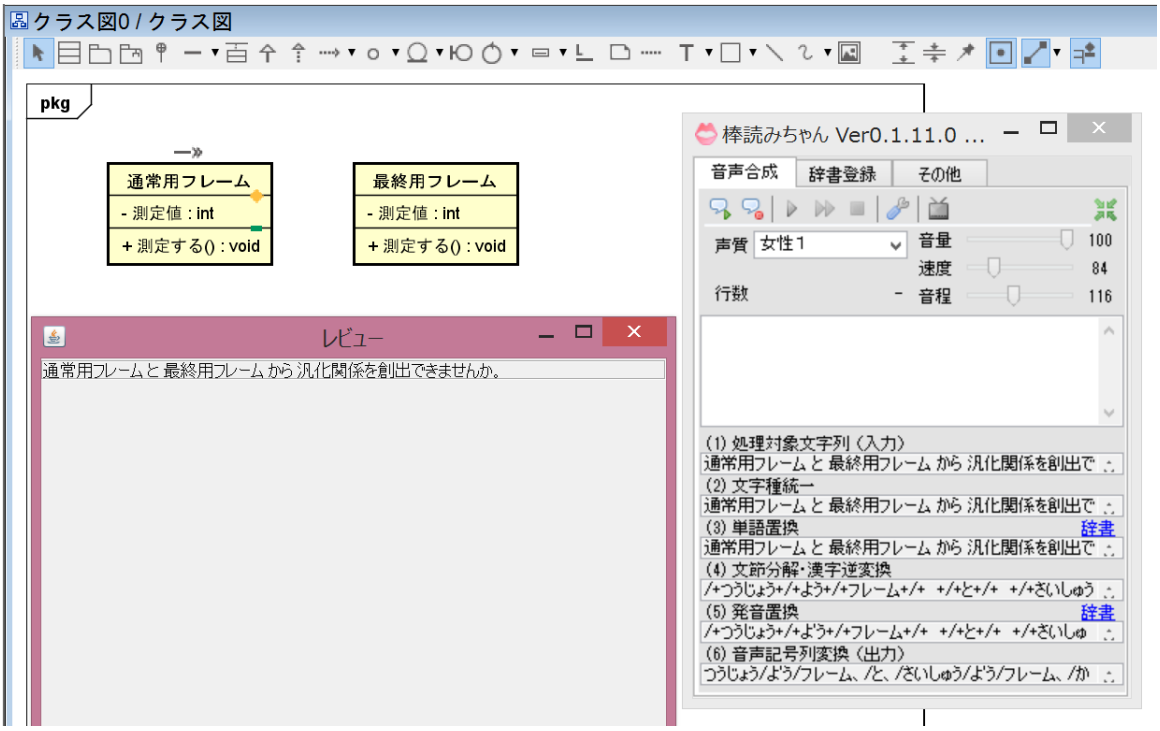


Figure 8: Review Results

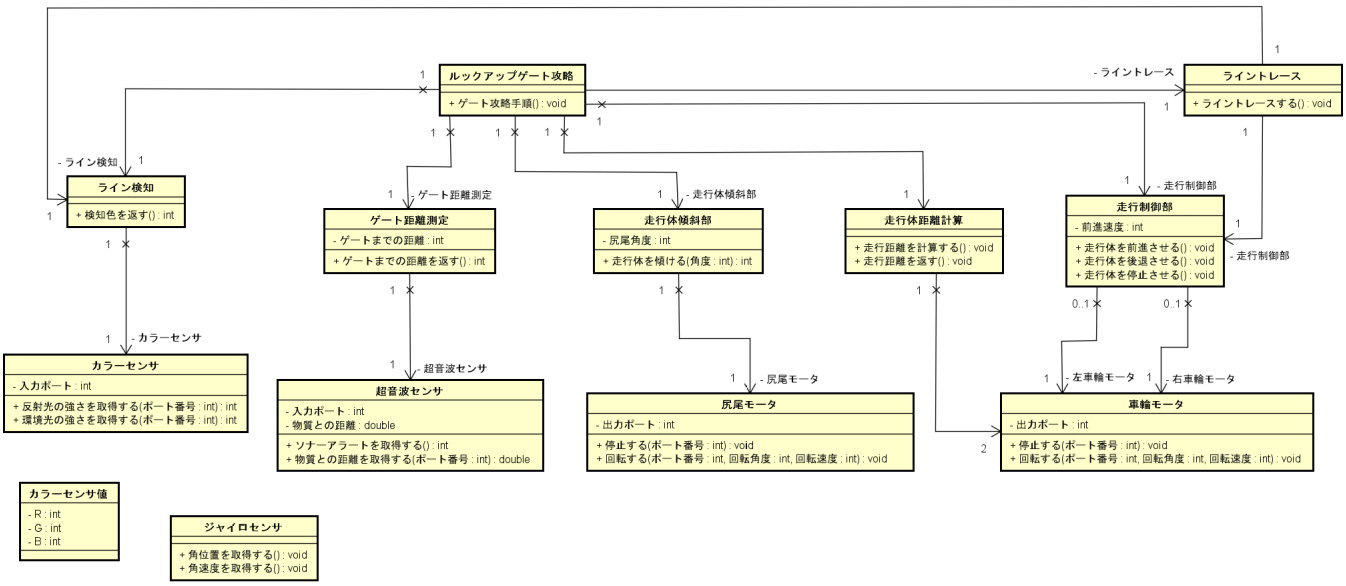


Figure 9: Class Diagram for RQ1

Therefore, we first calculate the similarity of class names between classes. If the similarity is equal to or greater than the threshold, the similarity for each attribute and operation in the class are calculated. Then, if the average of the elements is equal to or higher than the threshold value, it is possible to create a generalization relation from among the classes. Then a review that proposes creation of a generalization relation to the class diagram. If all classes have been checked, and a value above the threshold is not detected, a sentence notifying that there is no class to be reviewed is prepared in advance, and the sentence is sent to BouyomiChan.

4 EXPERIMENTS

4.1 The Goals

Experiments will be conducted to check if the review assistance system can perform useful reviews.

First, we check whether the review assistance system performs a series of actions proposed in the prior paper[16].

Secondly, we investigate whether it is possible to obtain useful review results by applying the review assistance system to multiple class diagrams actually created.

If useful results are obtained for both experiments, the results are considered to indicate practicality of the review assistance system.

Based on the above considerations, we set the research questions as follows.

RQ1 From the class diagram used in the evaluation study of the prior study[16], can we conduct a review similar to the prior study?

RQ2 Can useful reviews be obtained from multiple class diagrams actually created?

4.2 Evaluation Experiment 1

This experiment is for RQ1.

4.2.1 Protocol

The class diagram used in the preliminary evaluation experiment is shown in Figure 9. The material is created by a team participating to ET Robocon 2017[17]. the preliminary evaluation experiment has shown that it is possible to propose the creation of generalization relationships from the "tail motor" class and the "wheel motor" class in this class diagram. We apply the review assistance system to the class diagram and see if the same result as this result can be obtained.

4.2.2 Results

Figure 10 shows the result of applying the review assistance system to the class diagram in the Figure9. The pop-up display shows the results similar to those in the previous study[16].

4.3 Evaluation Experiment 2

This experiment is for RQ2.

4.3.1 Protocol

The product of the team participating in ET Robocon 2018 [18] is applied to the review assistance system. Multiple class diagrams created at the UML learning stage. We check whether a useful review can be obtained. Figures 11 and 12 show the class diagrams used in the experiment.

It is desired that generalization relations are suggested from three classes "inclination control," "inverted control," and "tail control," and from two classes "gyro sensor" and "color sensor" in Figures 11 and 12, respectively.

4.3.2 Results

The results obtained by applying to the class diagrams in Figure 11 are shown in Figures 13 and 14. The threshold represents filtering threshold for the similarity of attributes and operations.

At a threshold of 0.35, we can propose the creation of a generalization relation from "inclination control," "inverted control" and "tail control." Although "line trace running" and "running control" already have a generalization relationship, they include common operations and are included in the review because they are inappropriate to describe the generalization relationship. If the threshold is lower than 0.32, we have made a proposal from classes that do not require the creation of a generalization relationship.

The results of applying the review assistance system to the class diagrams in Figure 12 are shown in Figures 15, 16, and 17.

At a threshold of 0.35, we could not propose the creation of a generalization relationship. By lowering the threshold to 0.32, we could propose the creation of a generalization relation from "inverted control" and "tail control." When the threshold was lower than 0.30, a proposal was made from classes that do not require the creation of a generalization relationship.

It was not possible to propose the creation of a generalization relation from "gyro sensor" and "color sensor" for any threshold in either class diagram.

4.4 Discussion

In Evaluation Experiment 1, results similar to those of the preliminary experiment, were obtained. Therefore, we can say that the proposed method is correctly reflected in the review assistance system.

In the evaluation experiment 2, although some useful reviews were obtained, the result changed depending on the threshold. Since whether a useful review can be obtained depends on threshold, depending on the class diagram, effective approach will be to automatically derive a useful threshold for the target class diagram or to change the threshold using parameters.

Since "gyro sensor" and "color sensor" are the unknown word of Kuromoji, it cannot divide into morphemes. This is the reason that we cannot propose creation of generalization relation from "gyro sensor" and "color sensor." It is possible to register these words in the dictionary of Kuromoji.

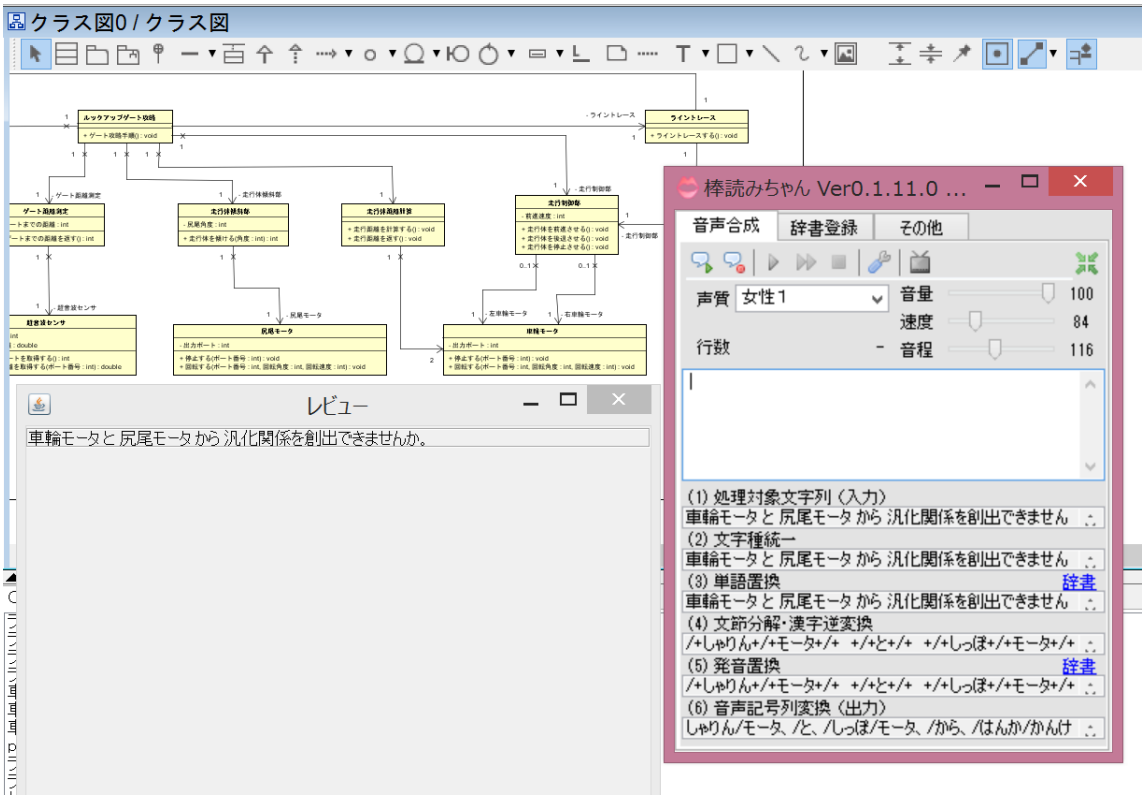


Figure 10: Review Result of RQ1

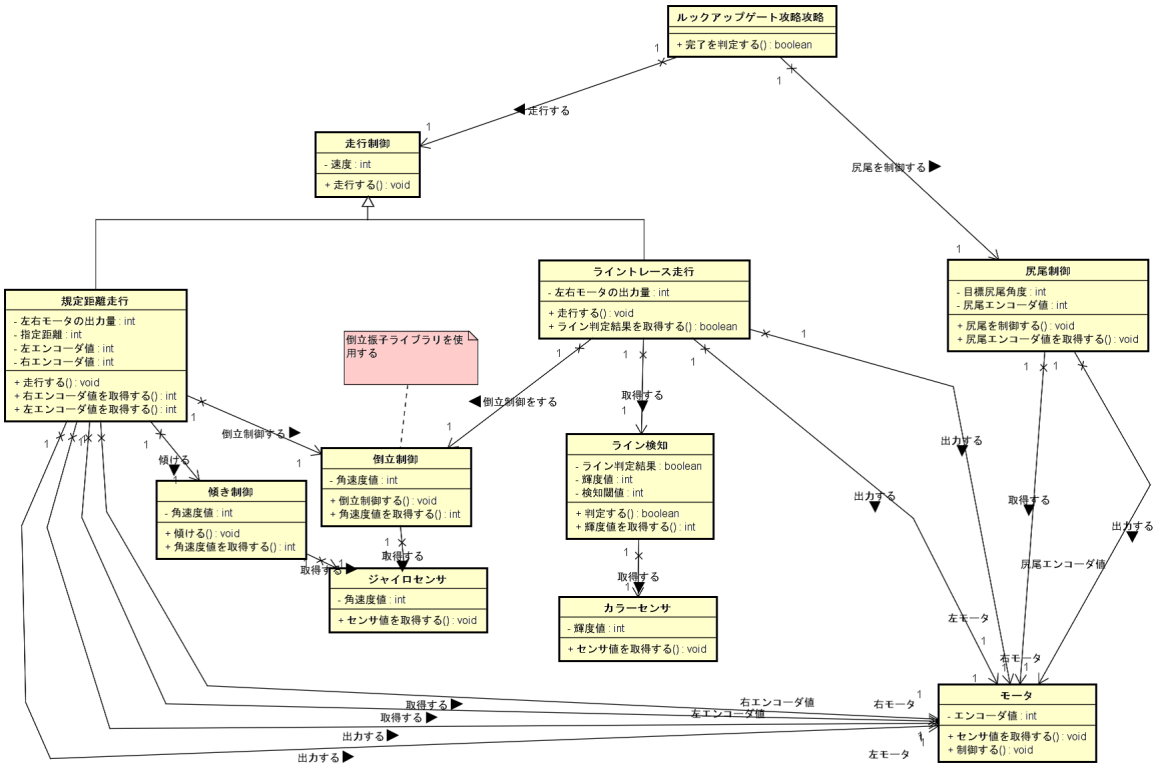


Figure 11: Set 1 of Classes for RQ2

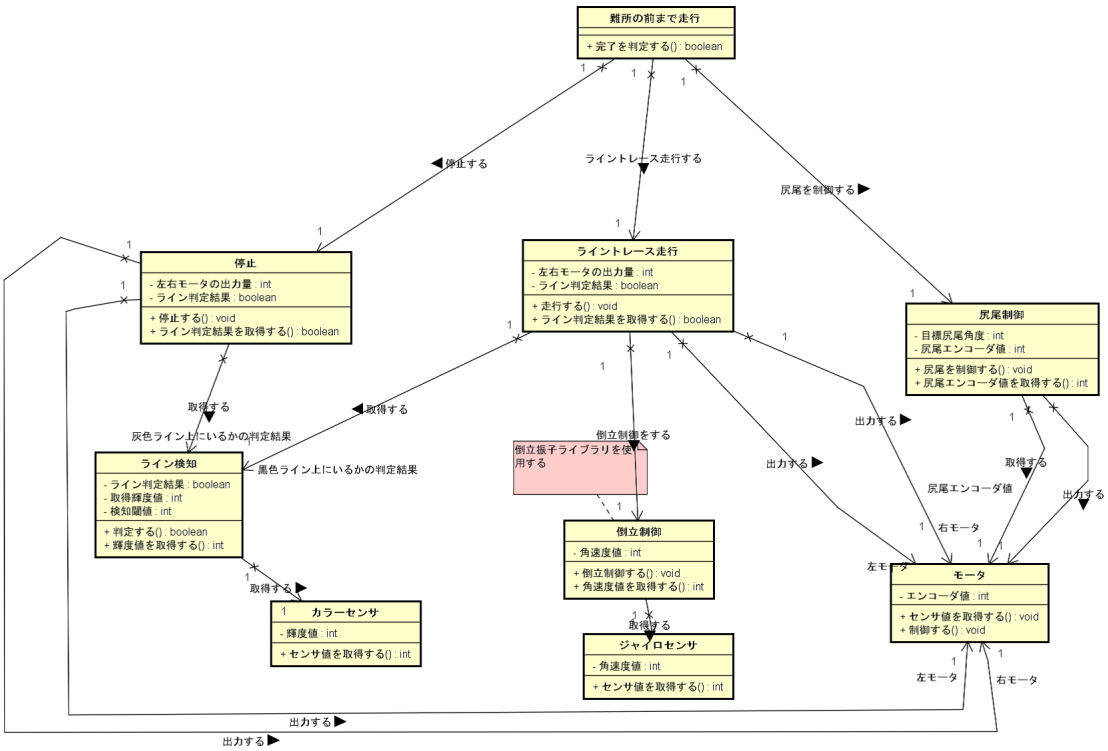


Figure 12: Set 2 of Classes for RQ2

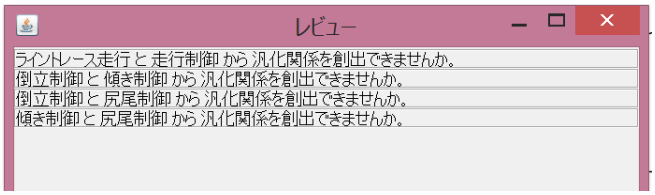


Figure 13: Review Result for Fig.11, Threshold=0.35

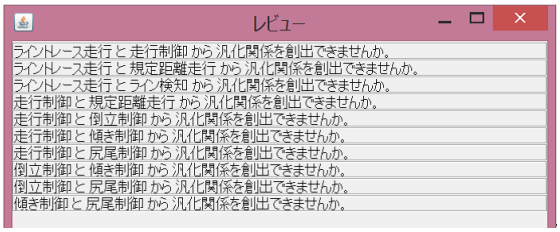


Figure 14: Review Result for Fig.11, Threshold=0.32

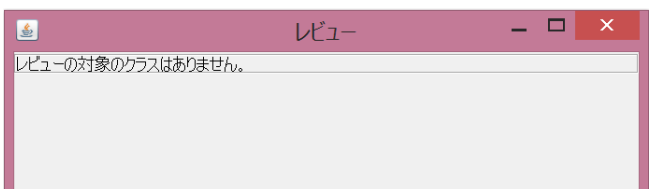


Figure 15: Review for Fig.12, Threshold=0.35

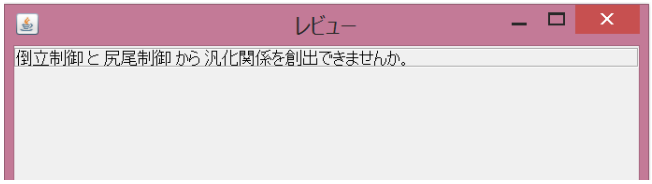


Figure 16: Review for Fig.12, Threshold=0.32

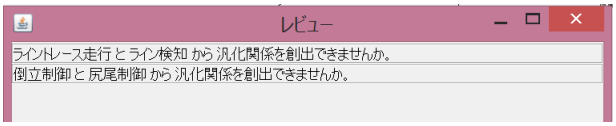


Figure 17: Review for Fig.12, Threshold=0.30

This time, evaluation experiments are not the cases, however if the class name is a single word, then the Jaccard coefficient for calculating the similarity of sets cannot measure the similarity. It leads un-useful review results.

For both of the experiments, we cannot create inclusion relations. The target class diagrams do not have such possibilities even when we check them manually.

5 RELATED WORK

Nakamura *et al.* [19] measures the comprehension of class diagrams from the word similarity used in class diagrams. Based on the hypothesis that words with a high degree of similarity, that is, words with strong word relationships, are highly comprehensible when placed close on a class diagram, the paper proposes a measurement method for understandability of class diagrams

In our work, the class diagram is reviewed based on the degree of similarity used in the class diagram. It may be effective to inspect the class diagram using the degree of similarity as [19].

6 CONCLUSION

In this paper, we developed a review assistance system using the method proposed in the previous research. It is able to propose the creation of a generalization relation among classes with a certain degree of similarity.

Future study includes the implementation of similarity calculation methods other than Jaccard coefficients, such as calculation of semantic similarity, and the setting of threshold values. To perform experiments on examples including both of generalization relations and inclusion relation is also included in the future study.

ACKNOWLEGEMENT

The research is being partially conducted as Grant-in- Aid for Scientific Research C (16K00094) and C (17K00111).

REFERENCES

[1] Object Management Group: "About the Unified Modeling Language Specification Version 2.5.1," https://www.omg.org/spec/UML/2.5.1/, (accessed 2019/05/20)

[2] Reiko Matsukawa, Mika Kouda, Suzuka Muramatsu, and Kanichirou Wakayama: "On Difference among Learning Effectiveness Proposed by Visual, Audio, and Text, (title is translated by author, original title is in Japanese)" Japan Society of Educational Information, Vol.9, pp.120-121 (1993) (In Japanese)

[3] Ruth Clark: "Evidence-Based Training Methods —A Guide for Training Professionals—, 2nd Edition," ATD press (2014)

[4] Eric Evans: "Domain-Driven Design: Tackling Complexity in the Heart of Software, "Addison-Wesley Professional (2003)

[5] Toyotaka Hagiwara: "Safety and Security sekkei no tameno Modular Approach to Jun-Keishikishuhou no sekkei genba heno Tounyuu —Software no 5s3 tei de kekkan no nai sofutware wo tsukuru [Introduce of Modular Approach and Semi-Formal Method to Design Site for Safety and Security — Building Reliable Software based on '5s3' on Software]," Report on IPA `https://www.ipa.go.jp/files/000057617.pdf` accesed 19th July 2019

[6] astah*, http://astah.change-vision.com/ja/ (accessed 2018/06/25)

[7] Shinji Sako, Chiyomi Miyajima, Keiich Tokuda, and Tadashi Kitamura: "A Singing Voice Synthesis System Based on Hidden Markov Model," IPSJ Journal, Vol.45, No.3, pp.719-727 (2004) (In Japanese)

[8] Hideki Kenmochi, Hayato Ohshita: "Singing synthesis system 'VOCALOID' Current situation and todo lists,", The Special Interest Group Technical Reports of IPSJ MUS, Vol.2008-MUS-074, pp.51-56 (2008) (In Japanese)

[9] Graham Wilcok: "WikiTalk: A Spoken Wikipedia-based Open-Domain Knowledge Access System," International Conference on Computational Linguistics (Mumbai 2012), Proceedings of the Workshop on Question Answering for Complex Domains, pp.57-70 (2012)

[10] BouyomiChan, http://chi.usamimi.info/ Program/Application/BouyomiChan/ (accessed 2018/10/24).

[11] Fritz Zwicky: "Discovery, Invention, Research, through the Morphological Approach," Science, 163 (3873) (1966)

[12] Kuromoji-Atilika, https://www.atilika.com/ja/kuromoji/ (accessed 2018/06/25)

[13] Suphakit Niwattanakul, Jatsada Singthongchai, Ekkachai Naenudorn, Supachanun Wanapu: "Using of Coefficient for Keywords Similarity," International Multi-Conference of Engineers and Computer Scientists (Hong Kong 2013), pp.380-384 (2013)

[14] Martin Flower: "Refactoring: Improving the Design of Existing Code, 2nd Edition," Addison-Wesley (2018)

[15] astah* API, http://members.change-vision.com/javadoc/astah-api/8_1_0/api/ja/doc/index.html (accessed 2019/02/20)

[16] Kozo Okano, Kazuma Takahashi, Yusuke Naka, Shinpei Ogata, and Toshifusa Sekizawa: "Analysis of Specification in Japanese using Natural Language Processing and Review Supporting with Speech Synthesis," IEICE Technical Report, vol. 117, no. 465, KBSE2017-52, pp. 79-84 (2018) (In Japanese)

[17] ET Robocon 2017, http://www.etrobo.jp/2017/ (accessed 2018/01/29)

[18] ET Robocon 2018, http://www.etrobo.jp/2018/ (accessed 2019/01/29)

[19] Yuto Nakamura, Hironori Washizaki, Yoshiaki Fukazawa: "Evaluation of Understandability of UML Class Diagrams," Proceedings of Software Engineering Symposium 2009, pp.165-172 (2009) (In Japanese)

Effectiveness for Continuous Using of Prompting Changeover System from Repetitive Behavior

Yuya Miyawaki†, Katsuhiko Kaji†

†Aichi Institute of Technology
1247 Yachigusa, Yakusa, Toyota, Aichi, 470-0392 Japan
b18726bb@aitech.ac.jp

Abstract - Children with developmental disorder have more difficulty engaging in changeover behavior than children without such disorders. Since this characteristic will be more critical as they grow and become adults, prompting changeover behavior is very important. When supporting prompting changeover, reducing the load on them and their supporters is also required. We developed support equipment and verified its effectiveness to provide assistance that prompts changeover by voice and vibrations. The changeover support is done based on the target object of the repetitive behavior. Therefore, since children with developmental disorders feel that the target object of the repetitive behavior is itself providing the support, they will respond to the prompt of the changeover support.

Keywords: Children with developmental disorder, Repetitive behavior, Changeover support, Support device

1 Introduction

Children with developmental disorders have more difficulty doing changeover behavior than children without them [1]. This is a major issue for children with developmental disabilities as they become adults and join society.The load on supporters during changeover support is another major issue.Based on this background, we reduced the changeover support load for both the benefit of children with developmental disabilities and their supporters.

This research proposes a device and support method that provides changeover support for developmental disabled children by voice and vibrations (Fig. 1). We designed a support device and implemented a prototype to achieve our proposed support. The prototype provided three kinds of support: vibration, speech, and operation. First, vibration support gives a moving stimulation through the target object that is held or worn by the target child. Second, speech support is generated from the target object and provides two types of changeover voice support: preparatory and parental voices. Third, operation support is provided so that the target child can understand the changeover support mechanism. To that end, the supporter directly manipulates the target object to overwrite the existing operation on the target object.

We verified our system to determine whether the prototype provides support. A family with a special needs child cooperated in our study and we video-recorded the child's performance with the supporter who also evaluated our system.

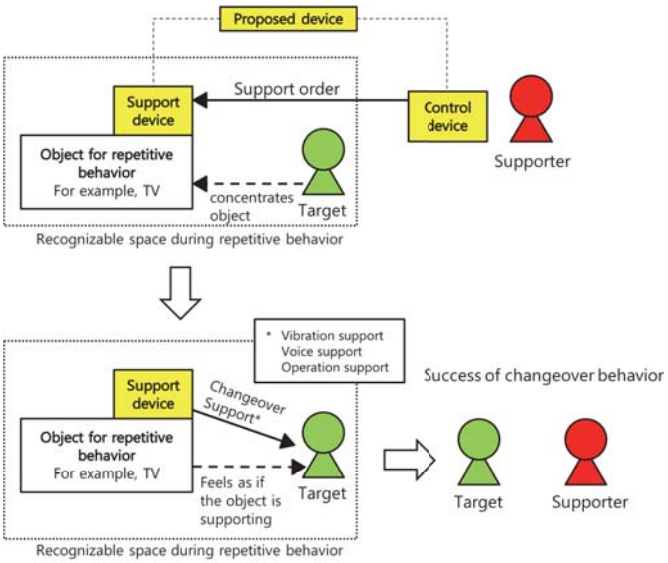


Figure 1: Support system outline

2 Related work

In recent years, research on developmental disorders has increased [2]. Although changeover support away from obsessive behavior is one basic type of support [3], unfortunately, few studies have addressed it.

However, if the changeover support does not work very well, the burden falls on the supporter [4]. In addition, smooth behavior in a group might be hindered. Therefore, in the research on changeover support, effective methods that provide general direct support and ways of creating ideal environments for changeover support have been discussed [5]: in other words, making successful prompts. Little research has focused on "what to do."

Support exists that matches the environment to the target child [6]. For example, the child ' s supporter waits until the environment and the target child ' s mindset can be switched. In this way, the supporter acts in coordination with the target child who can live based on his or her own ideal situation. Therefore, such support benefits the target child. On the contrary, forced changeover behavior increases the stress on the target child and strengthens the attachment to the obsessive behavior.

Voice support establishes changeover behavior by a repeated voice, for example, providing preliminary preparation for an imminent changeover. Voice support does not just denote that such sound is required. Timing and voice strength are also

important.

For digital devices, systems have been proposed that extend conventional support using tablets and PCs [7]. Using digital devices can significantly reduce the cost of producing individual tools, which is a problem with conventional support. This step will also reduce the burden on supporters. Unfortunately, many target children can't use digital devices because they have difficulty correctly inputting commands or they don't know which icons to touch.

3 Repetitive behavior for childlen with develomental disorder

The repetitive behaviors of children with developmental disorders are significantly different from those of children without them, especially in the time, period, and attachment force that occur. Children with developmental disorders show a marked attachment to certain items or situations. They never want to relinquish them, which causes repetitive behavior.

Repetitive behavior is associated with place. However, the target child only changes the repetitive behavior based on the place. In other words, the essence of repetitive behavior does not change.

Since the changeover from repetitive behavior is stressful for children with developmental disorders, they have difficulty changing. One major obstacle has been identified in the performance of children with developmental disorders and those without them. In addition, the group life of children with developmental disabilities is extremely difficult due to their repetitive behavior. Therefore, since they can't cope with their environment because it fails to match their feelings, they can't digest sudden schedule changes. Compulsory changeover to adjust to the surrounding situation causes a great deal of stress on the target child. As a result, he may engage in self-harm or hurt others. Forced changeover supports counterproductive effects.

However, children with developmental disorders only concentrate on objects of interest. Therefore, often support from the surrounding people is not transmitted. In addition, attachment to a supporter may occur. Such attachment to a supporter might turn him/her into a favorite, and they decide they prefer support from that supporter and not from the others. Hence, they only accept support from their favorite supporter.

Our proposed system provides the following support, as shown in Fig. 1. The target child concentrates on his target object. Perhaps the assistance from the surrounding supporters will not be accepted. In this research, the target child's recognition range is defined as the Repetitive Behavior Space. This space is conceptual and varies in scope and size depending on the target child. A support device is attached the child's target device to provide support to the objects of choice. The supporters benefit the target child using remote control equipment outside of the Repetitive Behavior Space. Our changeover promotion system generates voice and vibrations to support the target child who can recognize whether the target object itself is vibrating or speaking. This enables support from inside the Repetitive Behavior Space that is expected to encourage changeover behavior. Even if the support person changes, the voice support from the object of attention does not change. This resolves the issue of attachment to a particular supporter.

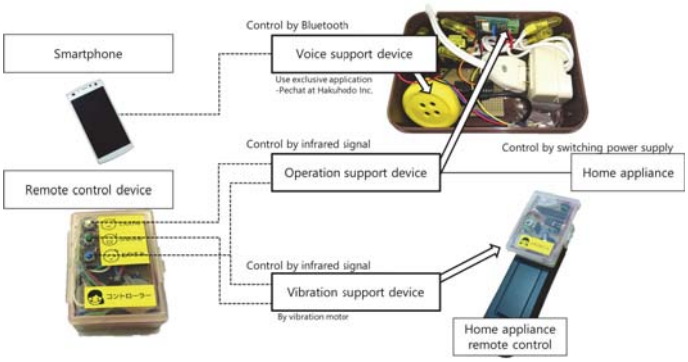


Figure 2: Support system detailed overview

The vibration support device generates vibrations on the target object of interest to inform the target child of the support. Since this support uses vibrations, the target child must be touching the target object to receive support. When not making contact, an object must be used that is in contact with an item that is related to repetitive behaviors.

The speech support system provides support using speech from the target object. The audio attached to the object of interest plays back a sound that can facilitate the changeover. We chose the same voice used at homes and educational settings based on our assumptions that it would be effective.

The operation support device enhances the effect of our proposed system by controlling the target object ' s power so that operational control cannot be relinquished to the target child. This device is only effective when the object of interest runs on electricity.

Three devices are controlled by a remote-control device, which wirelessly controls the device.

We conducted interview surveys and observed the home environments of the families from whom we obtained prior consent for cooperation with our research. Based on the information obtained from the household environment, the interview surveys, and home observations, we tailored our equipment to the living environment of the target child. Since target of the basic commitment at home was watching DVDs, we installed our equipment on a DVD and attached our equipment to it.

Based on a preliminary survey, the child's attention behavior was performed for the following period: 1) from waking up to going to school and 2) from returning home to taking a bath.

Changeover behavior is classified as complete or incomplete. Complete changeover is defined as changeover behavior that can't continue the repetitive behavior after the changeover.Such changeover behavior is completely separated from the target object. Incomplete changeover is defined as changeover behavior where the repetitive behavior can continue after changeover.

We attached a vibration support device to the remote control so that it could always be used during playback. We attached a small device to the remote control to reduce the risk

Figure 3: Vibration support device

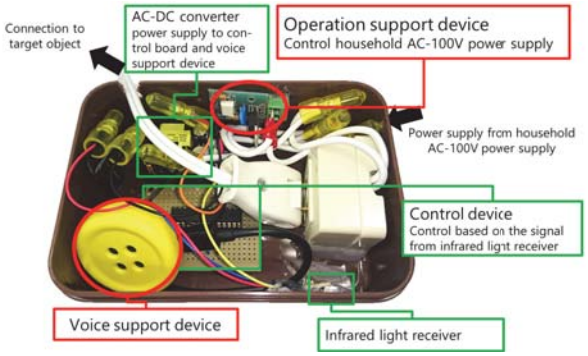


Figure 4: Integrated voice and operation support devices

of inconvenience when using it. Fig. 3 shows the remote control to which we attached our mounted device.

The target child ' s repetitive behavior is watching DVDs. In this experiment, we defined the obsessed object as the object of repetitive behaviors. Therefore, the target object in this example is the TV screen. The voice support device needs to be installed around it so that the sound can be heard. The home appliance control support device needs to be installed around the output cable of the DVD player. We integrated the voice support device and the home appliance control support device next to the TV display. The target object ' s power is drawn to the device, and control is done using an infrared signal. The voice support device also incorporates an independent control mechanism using a commercially available system and controls from a smartphone application. Fig. 4 shows the attached device, and

The supporter operates each device at each changeover action. Basic operations are done using the infrared remote control that was implemented in an easy-to-understand manner using large operation instructions and colored buttons. By pressing the button, the infrared signal is emitted to control the target device. The voice support device can record, register new assistance, and so on. Each operation is performed using a smartphone application.

4 Verification of effect by support from the target object

Our proposed system provided continuous support with the cooperation of a female junior high school special support school, a first grader, and her family. CWhen the child is engaged in repetitive behavior, she tends to prioritize the repetitive behavior over the voices of the surrounding supporters. Therefore, in many situations she cannot transition to the changeover action.

Table 1: Classification of changeover behavior in child

	Complete changeover	Incomplete changeover
Favorable		Message
Unfriendly	Meal Going out Doing homework Taking a bath	Brushing teeth
Voluntary	Going to the bathroom	

We recorded from March to September 2018. The recording was mainly done by the mother who was also a supporter at home, and we did video recording on our visitation days. We observed the changeover behaviors that occurred at home and classified them with the supporters. The classifications are shown in Table 1. We predicted and classified the situations we observed and determined whether the target child liked or disliked the changeover behaviors itself.

As shown in Table 1, we made two major classifications based on the content of the changeover behavior. Complete changeover behavior is defined as changeover behavior that cannot be continued after the changeover. Since repetitive behavior can ' t be continued after a changeover, we can't changeover a behavior even with such support such as the direct voice of supporters and gestures. Incomplete changeover behavior is defined as repetitive behavior that can continue after a changeover. In many cases, changeover behavior can also be done by summoning the supporters.

We propose support based on the timing of the occurrence of the daily changeover behaviors. The supporter operates the support device based on the procedure. The supporter observes and records the target child ' s behavior.

Commitment behavior in the home is performed intermittently from the time of getting up for school on weekdays, going out (weekends and holidays), and returning home for a bath. The target child watches his favorite DVD.In some cases, he watched on a mobile DVD player.

In this experiment, we calculated the success rate of the prompting of the changeover behavior using our proposed support. Then we analyzed its effectiveness. The success rate was based on situations where the target child did the changeover behavior just based on reminders from the support device. First, the support person operated the support device for the changeover. Next the support person called out to promote the changeover behavior. If no changeover behavior is seen in this state, the strength of the support from the support person is sequentially increased to promote changeover behaviors. We increased the strength of the support in the following order: direct support from the supporter, gestures, stopping the object of interest, and taking arms and guiding.

We recorded the extent of the increase of the support intensity in a five-step evaluation. In addition to the changeover action results, the target child ' s daily situation was recorded and analyzed.

The graph obtained by aggregating the weekly changeover action success rate based on our proposed support is shown

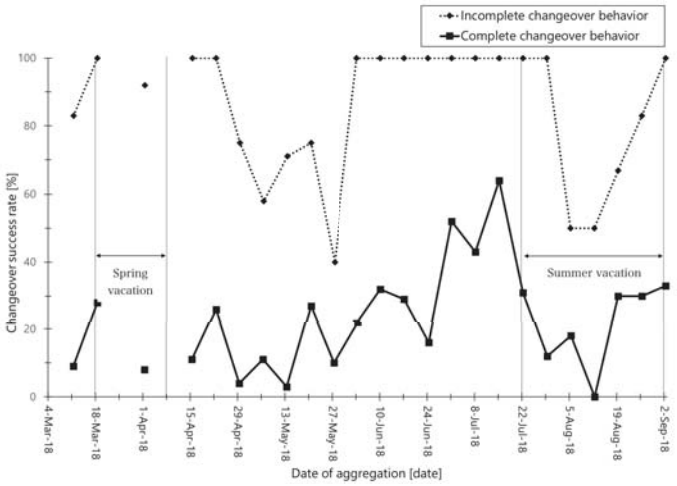


Figure 5: Graph of changeover action success rate

in Fig. 5. The complete changeover behaviors can't continue the repetitive behavior after the changeovers. Therefore, the changeover success rate is low. On the other hand, incomplete changeover behavior can continue the repetitive behavior after the changeover. So the target child easily made the changeover.

We focused on such changes in the physical condition and the environment and analyzed their occurrence rates, the changes in the environment, and the changes in the continuation of support.

The success rates of the changeover behavior between the complete and incomplete changeover behaviors changed depending on the physical condition of the target child. In the week where the changeover behavior success rate decreased, both the complete and the incomplete changeover behaviors decreased, and the rate increased in subsequent weeks.

In the incomplete changeover behavior, almost 100% was maintained in the week when the child had a good physical condition and a good environment. Therefore, for the incomplete changeover behavior, we continue to propose support and establish the changeover behavior.

A major change in the environment revolved around whether to go to school. The experimental period included spring vacation from March 18 to April 8 and summer vacation from July 22 to September 2. We confirmed a decrease in the changeover behavior success rate during both holidays.

Although the changeover behavior success rate fluctuates based on environmental changes and physical conditions, the appearance tendency tends to improve as a whole. This improvement trend shows better use of our continuous proposed support. The decrease in the success rate due to environmental and physical condition factors can be judged by a comparison with the incomplete changeover behavior, which is the success rate at a normal 100%. In other words, except for the week when the changeover success rate of the incomplete changeover behavior decreased, the general tendency showed an increase. The assistance with our continuous proposed support is effective for successful changeover actions.

5 Conclusion

We reduced the burden on both the target child and the supporters during changeover behavior from obsessive behavior and obtained spontaneous changeover behavior. Continuous support was maintained, and we eventually achieved an emergence rate of 60%, proving that this support is effective. On the other hand, a 100% success rate can be maintained in incomplete changeover behaviors. We believe that our process has advanced to a practical stage.

In the future, we will provide more continuous support and investigate how long it can improve the success rate in accordance with environmental changes and physical conditions. We must also analyze the problems of our proposed support based on the situation and changes of the target child and investigate whether a significant adverse effect exists.

We received the following comments from a supporter. "Using the device as a trigger for changeover behavior allowed smooth changeover behavior." "Although at times using the device was burdensome, it switched my feelings." "The child asked to use the device." In other words, using the device itself is a burden. However, the support environment was improved by the device.

REFERENCES

[1] Fred R Volkmar, Rhea Paul, Ami Klin, and Donald J Cohen. *Handbook of autism and pervasive developmental disorders, diagnosis, development, neurobiology, and behavior*. John Wiley & Sons, 2005.

[2] Tristram Smith, Annette D Groen, and Jacqueline W Wynn. Randomized trial of intensive early intervention for children with pervasive developmental disorder. *American Journal on Mental Retardation*, 105(4):269–285, 2000.

[3] O Ivar Lovaas. *Teaching individuals with developmental delays: Basic intervention techniques.* Pro-ed, 2003.

[4] Jo Bromley, Dougal Julian Hare, Kerry Davison, and Eric Emerson. Mothers supporting children with autistic spectrum disorders: Social support, mental health status and satisfaction with services. *Autism*, 8(4):409–423, 2004.

[5] Dilek Erbas. A collaborative approach to implement positive behavior support plans for children with problem behaviors: A comparison of consultation versus consultation and feedback approach. 45(1):94–106, 2010.

[6] Rose Iovannone, Glen Dunlap, Heather Huber, and Don Kincaid. Effective educational practices for students with autism spectrum disorders. *Focus on autism and other developmental disabilities*, 18(3):150–165, 2003.

[7] Muhamad Fairus Kamaruzaman, Nurdalilah Mohd Rani, Harrinni Md Nor, and Mustafa Halabi Haji Azahari. Developing user interface design application for children with autism. *Procedia-Social and Behavioral Sciences*, 217:887–894, 2016.

A Life Watching Service for Elderly People by using Interactive Home Robot

Keiichi Abe, Hiroki Fuse, Makiko Okumura, Masao Isshiki

Kanagawa Institute of Technology, Atsugi-City, Kanagawa, Japan
abe@he.kanagawa-it.ac.jp

Abstract – In Japan, the number of elderly people living alone is increasing with the declining birthrate and aging population. Under such circumstances, life watching services for the elderly are attracting public attention. In this paper examined whether the robot could be applied as a new life watching service by analyzing a log recording the talks between the watching target and the robot. We developed and evaluated a trial system of life watching service using an interactive home robot proposed in this paper.
Keywords: Communication Robot, Life Watching Service, ECHONET-Lite.

1. INTRODUCTION

In Japan, the number of elderly people living alone is increasing with the declining birthrate and aging population. Under such circumstances, life watching services for the elderly [1, 2] are attracting public attention. By the way, there are many services [3, 4] that use robots for the watching. The basic functions of the existing watch robots are communication, and various functions such as telephone functions and schedule functions are implemented. Many of these robots use cameras for the watching. A robot equipped with a camera [3, 4] can watch the life of the target person with video, but it can not watch the target person when he/she is in a room where the robot is not installed or within a blind spot of the camera. Also, the subjects may have multiple problems such as the mental stress that they are always taken with a camera [5].

Therefore, in this study, we propose a method to monitor their life without feeling the mental stress by utilizing an interactive home robot for watching the life of the target person such as the elderly who live apart from the family.. Also, the dialogue commission type home robot for life watching that we developed this time starts talking when someone is detected by the human sensor etc. The robot also produces facial expressions and movements such as "nodding" and "swinging" when talking. We will describe how such a method helps to increase the chances of conversation in an easy-to-talk manner. In this paper, Chapter 2 is an outline of life watching service using the interactive home robot proposed by this research, Chapter 3 is prototype development, Chapter 4 is evaluation by demonstration experiment, Chapter 5 is questionnaire evaluation, Chapter 6 presents the conclusion.

2. LIFE WATCHING SERVICE USING INTERACTIVE HOME ROBOT

Chapter 2 gives an overview of a life watching service using the interactive home robot (hereinafter referred to as "IHR") proposed by this paper and details of the prototype IHR developed this time.

2.1 Outline of life watching service using IHR

Figure 1 shows the outline of the life watching service using IHR proposed by this paper. This service has an implemented function to record the contents of conversations between IHR and users as a log. Log data is saved as a file in CSV format and analyzed by a control program of the interactive home robot and used for the life watching. IHR's natural conversation was realized by combining Google Speech API of speech recognition service and chat conversation API of NTT docomo [6]. Classification numbers are provided according to the contents of the log data acquired by IHR (conversation, home appliance operation, etc.), and IHR estimates the life behavior of elderly people living alone by graphing the data in time series etc. Using those functions, we devised a system service to monitor life by comparing the living condition data with those acquired in the past. The final analysis results of life surveillance are transmitted to families and relatives living in remote areas by e-mail. The IHR is designed for use in a smart house [7] where home appliances and sensors compatible with ECHONET-Lite are connected to the home network.

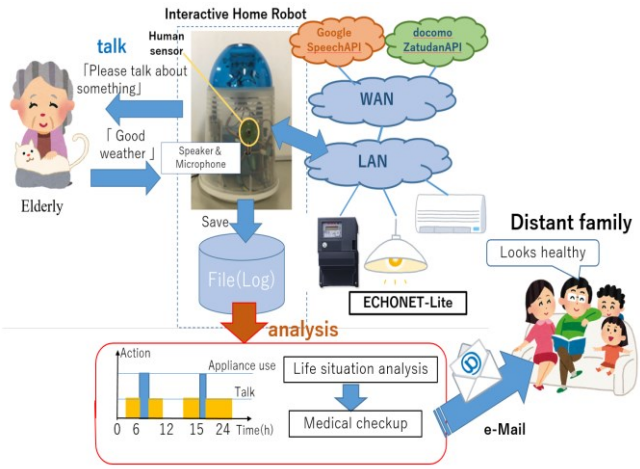


Figure 1: Overview of life watching service using IHR

3. PROTOTYPE DEVELOPMENT OF IHR

3.1 Hardware of IHR

In developing the IHR this time, we referred to the Magbot [8] devised by Professor Koike and modified it to enable natural conversation with the user freely. Figure 2 shows the system configuration of the IHR developed this time. As shown in Figure 2, RaspberryPi3 B + (hereinafter referred to

as RPI) is used as the main MCU (Micro Control Unit) of IHR, and distributed processing is performed using the Arduino microcomputer as the sub MCU. RPI is in charge of the IHR main processing, performs communication processing with ECHONET-Lite compatible home appliances, etc., performs speech synthesis processing and communicates with various network services (Google Speech API and docomo chat conversation API), etc. The Arduino microcontroller performs sub-controls such as the light emission of five LEDs that appear to be the face of the IHR and the swing motion of the robot by a servomotor. Therefore, the Ardunio microcontroller controls the part of the IHR facial expression and movement. In addition, a human sensor (Model Type EKMB1301111K / Panasonic) was installed on the body of the IHR to enable the IHR to actively talk to a user near the IHR. The sensor circuit was designed to react to a person within 20 cm of detection distance. Therefore, the proposed IHR operates only when a human movement is detected, and a program that can record the result as a log in the IHR file is implemented in the RPI. The main program of this IHR RPI was developed in Python 2.7 language, and the processing program on the Arudino microcomputer side was developed in C language.

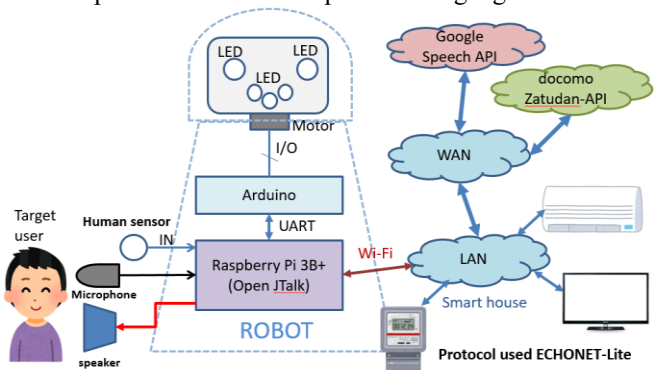


Figure 2: System of IHR

3.2 Mechanism of natural conversation

In order to allow the target person to talk naturally with IHR, we used Google Speech API for speech recognition service and the docomo's chat conversation API for natural conversation service for our IHR. Figure 3 shows the flow of the natural conversation processing implemented in IHR [6].

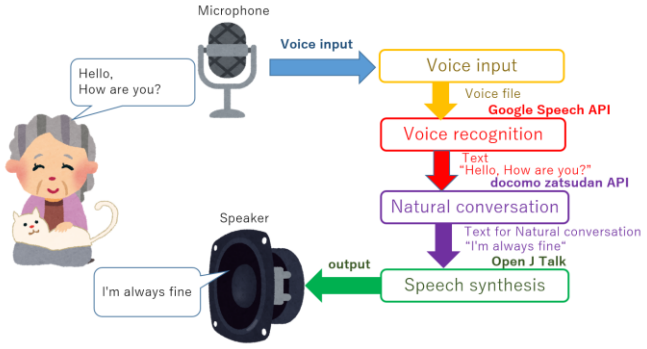


Figure 3: Flowchart of natural conversation processing of IHR

The IHR acquires the user's voice information with the microphone connected to the RPI, and first creates a voice input file. It performs speech recognition processing that converts speech input files into text information using the Google Speech API. The converted speech text information is sent to docomo's Zatudan-API, which creates text information containing natural speech information. This text information was speech-synthesized using Open JTalk prepared in IHR, and the obtained voice conversation is outputted from the speaker connected to RPI. In this manner, the natural conversation between human and IHR was realized.

3.3 IHR facial expression

Figure 4 shows the face of the IHR developed this time. The robot's face was made of a breadboard, five LEDs, and five resistors. We arranged two large-size LEDs side by side, protruding as the eyes, and three LEDs located below the center as the mouth. Various expressions are created by lighting and blinking each LED as shown in Fig. 4 as needed in the conversation. In practice, a translucent plastic mug was placed over the breadboard to prevent the circuit from being exposed. When talking with people, IHR lights its eyes and mouth, and produces a swinging motion with the DC servomotor so that it looks friendly and is easy to talk to.

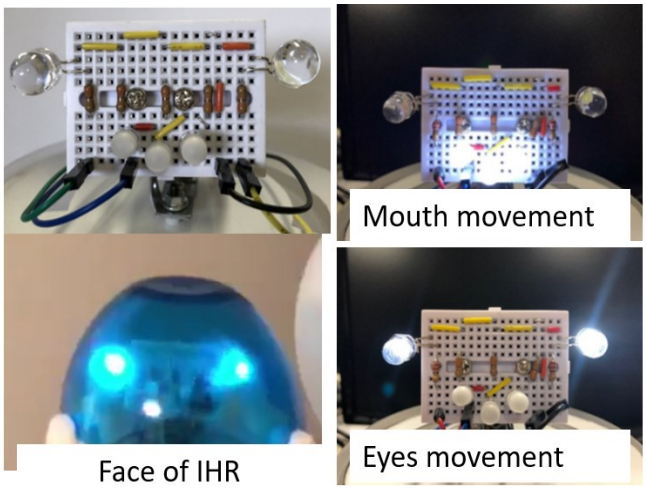


Figure 4: Facial expression of IHR

3.4 Conversation log recording function

IHR has a function to keep a log of the talks of IHR with the target person of the life watching. In order to distinguish between daily conversations and household appliance operation requests as the conversation content, classification numbers are provided to make it easy to extract necessary data using programs. The classification number is 1 for daily conversation, 2 for household appliance operation request, and 10 for no conversation. This "no conversation" is written as a log to a file inside IHR when nobody is detected by the human sensor and there was no conversation.

3.5 Life watching method by conversation log analysis

By analyzing the log data acquired by IHR using a program or application, the life rhythm and behavior of the watch target person can be estimated to some extent. For example, a set of log collected over a long period can be treated as reference data of the living behavior of the target person. Comparing the latest data with this reference data makes it possible to detect a continuation of an abnormal situation and notify the situation to relatives and families living in remote areas by e-mail etc. For example, if the state without conversation has continued for half a day or more during the time when such conversations are normally expected, it is possible to determine that the life behavior is different from usual, so IHR notifies the situation to the family and relatives by e-mail etc. with a message such as "Please call Mr. OO/Mrs.○○," to prompt them to contact the watching target. The e-mail transmission is performed by the main program of RPI, and realized by the standard library of python.

3.6 Acquisition of living information and appliance operation in the smart house

The IHR developed this time uses ECHONET-Lite [9] as a communication protocol for acquiring living information and operating home appliances in the smart house.

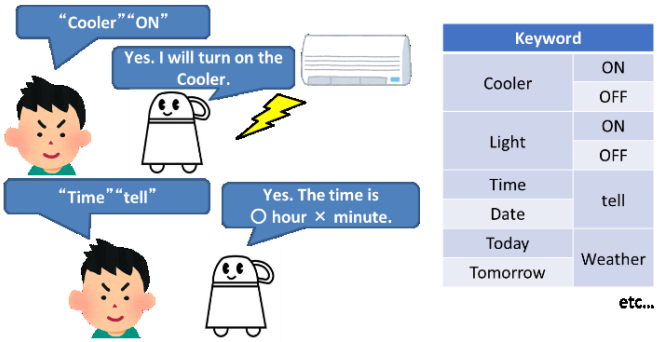


Figure 5: Appliance operation and information service by keyword of 2 words

By communicating with smart meters and home appliances in the smart house, IHR is able to acquire the power consumption of the entire home, acquire ON / OFF status of various home appliances, and remotely operate home appliances. Also, when the user asks time, weather, etc., IHR acquires necessary data in JSON format from the time [10] or weather web service site [11], and outputs it as voice from IHR. The IHR developed this time was designed to remotely turn on / off the home appliance with the words corresponding to the keyword or to provide the weather and time information when the corresponding keywords are included in the conversation. Figure 5 shows an example of the remote control of the home appliances and information provision of the weather and time by keywords. For example, if you ask the IHR to turn on the Cooler in the room, there is a possibility that the word "Cooler" may be misrecognized. Therefore, IHR is programmed to remotely turn on the "Cooler" only when the two words "Cooler" and "ON" have been recognized found. If you want to hear the weather information from IHR, IHR provides today's weather information by voice only when the two words "Today" and "Weather" have been recognized.

3.7 Software of IHR

Figure 6 shows the flow chart of the main software implemented in the prototype IHR.

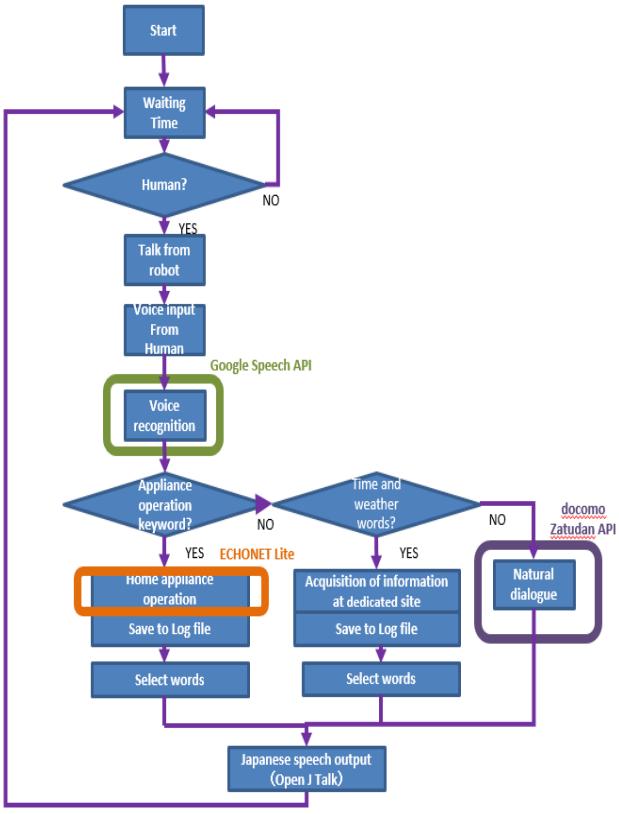


Figure 6: Main software flow chart of IHR

A pyroelectric infrared sensor was used as the human sensor. When the motion of a person is detected by the human sensor, the robot starts talking, and when the person is absent, the IHR does not speak. Upon receiving an input of human voice through the microphone, the IHR creates a voice file and sends it to Google Speech API to convert the audio information into text information. If there is no keyword related to home appliance operations in this text information, IHR checks whether there is a keyword related to time or weather. If there are no keywords related to time or weather, the text information is passed to docomo's chat service and converted to text information for natural conversation. The text information is converted to speech information using the speech synthesis software Open JTalk, and speech is output from the speaker to realize natural daily conversation between a human and a robot [6]. Also, if the 2-word keyword of the home appliance has been found recognized in the acquired voice information, the target home appliance compatible with ECHONET-Lite connected on the network is remotely operated and the result of the operation is presented to the user. Also, if there is a keyword related to time or weather, the IHR acquires information from the dedicated site and provides it to the user by voice.

4 EXPERIMENT

An experiment using the IHR prototype developed this time was conducted to demonstrate how accurately the living behavior of a subject can be analyzed from the log data acquired by the IHR.

4.1 Experimental method

This demonstration experiment was conducted at E602 on the 6th floor of Kanagawa Institute of Technology C2. The subject was a 20-year-old male, and the experiment period was from May 14, 2019 to May 20 except Saturday and Sunday. Figure 7 shows the system of the demonstration experiment, and Figure 8 shows the layout of the demonstration experiment.

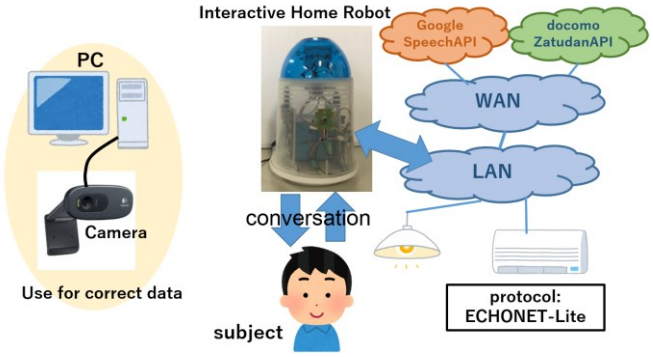


Figure 7: Experimental system

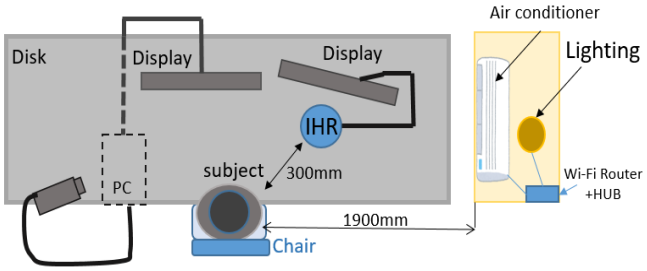


Figure 8: Layout of Experimental (Top View)

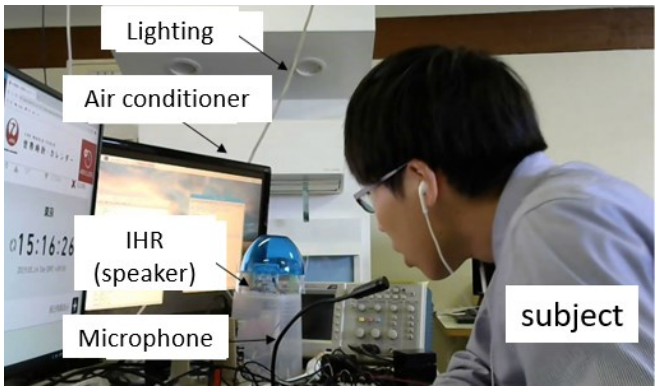


Figure 9: Demonstration experiment state from camera

Using the IHR prototype developed this time, we asked the subject to perform the usual behavior, and record information such as the subject talking with the IHR or remotely operating a home appliance via a robot as log data. The recorded information was analyzed to evaluate the accuracy of human behavior estimation. A camera was installed to acquire correct data in the demonstration experiment. Figure 9 shows a picture of the IHR and the subject during the demonstration experiment taken from the camera for acquiring the correct data. Table 1 shows the equipment used in the demonstration experiment. As shown in Figure 8, the two appliances used this time were an ECHONET-Lite compatible Cooler and lighting. The subject entered the room and had a conversation with the IHR at any time. During the demonstration experiment, the camera captured moving images of IHR and the user's conversation situation along the time display. The correct data was created manually while confirming the content and time of the action of the subject from the moving image of this camera. After the demonstration experiment, the reliability of IHR log data was evaluated by comparing the log data recorded by IHR with the correct data obtained from the camera's moving image. Also, as shown in Fig. 9, the subject sat at his desk and talked to the IHR only when necessary. The subject was asked to take action with a natural feeling, such as requesting the IHR to turn on the Cooler when it is hot, or to turn on the lighting when the room is dark. As shown in Fig. 8, the distance between the IHR and the subject was 300 mm, and the distance between the Cooler or lighting and the subject was 1900 mm.

Figure 10 shows log data acquired by IHR. Log data was saved as data in CSV format in order of date, time, conversation content, and classification number. This classification number is an optional item provided so that information can be easily extracted later by a database or Excel.

Tabel.1 Equipment used

Name	Type	Manufacturer
Microphone	MM-MCUSB25	SANWA-SUPPLY
Camera	C270	Logicool
Air conditioner	LDF7N-GX53/D2	TOSHIBA
Lighting	LDF7N-GX53/D2	TOSHIBA
speaker	SP-05	ALPEX

The classification number is 1 for conversation and 2 for home appliance operation and 10 for no conversation. In addition, the log data acquired by IHR is used to calculate the number of conversations and the number of operations of the home appliance in units of 30 minutes, and the accuracy rate of the number of conversations and the number of home appliance operations from comparison with the correct data was calculated by the following equations (4.1) and (4.2).

$$\text{Accuracy rate of conversation}(\%) = \frac{\text{Number of conversations captured in the log}}{\text{Number of conversations acquired by the camera}} \times 100 \quad (4.1)$$

$$\text{Accuracy rate of conversation}(\%) = \frac{\text{Number of appliance operations captured in the log}}{\text{Number of appliance operations acquired by the camera}} \times 100 \quad (4.2)$$

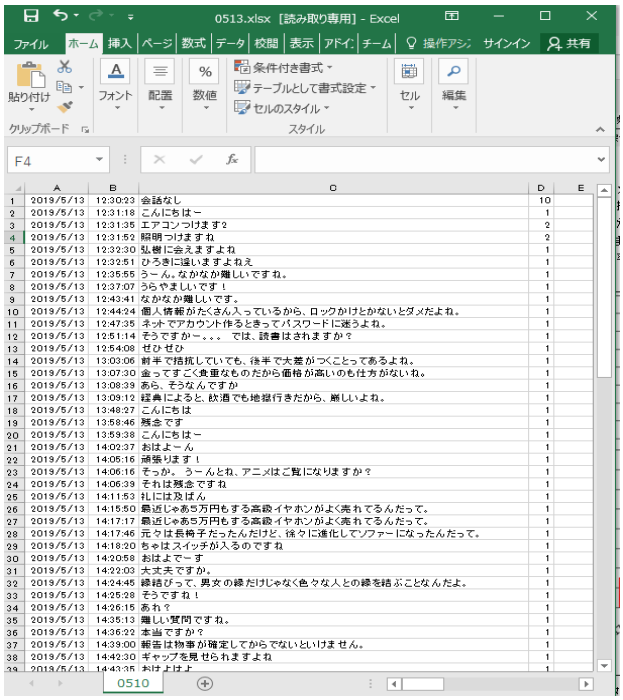


Figure10: Log data acquired by IHR (CSV format data)

4.2 Results and discussion

Figure 11 is a graph in which the log data obtained in the demonstration experiment from 14-May to 20-May, 2019 excluding Saturday and Sunday and the correct answer data acquired by the camera are summarized as the number of appliance operations and the number of conversations in 30 minutes. The horizontal axis is the time axis, and the vertical axis is the number of conversations and home appliance operations. The upper graph is the graph acquired by IHR, and the lower graph is the correct data obtained by the camera. In addition, Table 2 shows the correct answer rate of the number of conversations and the correct answer rate of the number of operations of the home appliance. As shown in Figure 11, the number of conversations in the log data acquired by IHR is larger than that of the correct data as a whole. As shown in Table 2, this resulted in the accuracy rate of the number of conversations exceeding about 100%. It is thought that this is because IHR responded by picking up the conversations of other persons because there were multiple people other than the subject during the demonstration experiment, and that the number of conversations increased. Next, referring to the 5-day data in Figure 11 in chronological order, the time zone for conversation, the type of conversation, etc. are almost synchronized with the correct data taken with the camera. Therefore, IHR's log data can be used to estimate the subject's behavior as well as the camera's correct data. The working hours of the subjects this time are 11:00 to 16:00, but it is possible to estimate the early departure and overtime hours etc. depending on the day of the week by combining with the subjects' action memos in Figure 11. It will be also possible to accurately estimate the time when the subject left for lunch. With 14-May and 16-May data, it is possible to clearly read the time spent leaving for lunch because there is no conversation with the IHR or the operation of home appliances, but with 15-May data it can not be judged whether the subjected was absent, because there is a conversation recorded. As we summarized the data in units of 30 minutes this time, an error of ± 30 minutes may occur depending on the time zone as compared with the correct data as shown in the graph of 17-May in Figure 11. For this reason, it is difficult to estimate the exact leaving time etc. when the subject leaves for a short time at lunch etc. in the analysis in units of 30 minutes. Table 3 compares the actual working hours with the working hours estimated from the log data acquired by IHR. In this result, it is within about ± 15 minutes, but it can be expected that the error will deviate by about 30 minutes if the room entry time is before 11:00, for example. Therefore, in log data acquired by IHR, it is possible to estimate living activities such as entering, leaving and working hours within an error range of about ± 30 minutes. In addition to this, since it is possible to determine whether the conversation content with IHR is a daily conversation or an operation of a home appliance, it is considered that it is possible to extract a subject-specific life behavior pattern from these kinds of information.

Table.2 Comparison of accuracy rate of log data vs camera (correct data)

Date	accuracy rate of conversation	accuracy rate of appliance operation
14-May-19	122%	100%
15-May-19	151%	93%
16-May-19	131%	100%
17-May-19	182%	100%
20-May-19	161%	100%

Table.3 Comparison of actual working time vs working time obtained from log data

Date	Actual working time	Working time obtained from log data
14-May-19	11:12～15:38	11:00～15:59
15-May-19	11:12～16:16	11:00～16:29
16-May-19	11:06～15:53	11:00～15:59
17-May-19	11:10～15:20	11:00～15:29
20-May-19	11:08～15:41	11:00～15:59

Figure 11: Number of conversation and appliance operation at from 14-May-2019 to 20-May-2019 (The upper graph shows IHR log data. The lower graph shows camera data)

5 CONCLUSION

The IHR proposed by this research is a technology that enables natural conversation based on voice information from users and voice control of home appliances. Although there is also a method using AI speakers as a related technology of voice conversation, most of the commercially available AI speakers have an inorganic design, so it seems that there are a certain number of people who find it difficult to talk to. Therefore, as a means for increasing opportunities for conversations with users, we developed an IHR that can express richly non-verbal information such as facial expressions and movements, but also human speech. By storing the contents of the conversation as log files and analyzing it, we examined whether it could be used as a service to notify the result of the living situation by e-mail etc. to distant family and relatives, that is, it could be used as life watching service. The prototype of IHR was developed this time, and the reliability of the data acquired by the log file was evaluated by comparing the log file acquired by IHR with the result of the moving image acquired by the camera. As a result, it became clear that when there were multiple people other than the subjects, the voices of other people were picked up, and the number of conversations counted in the log data of IHR was larger than the actual number. However, it was found that it was possible to estimate the behavior at which time the subject talks with the IHR well and operates the appliance. Therefore, in this study, in the log data acquired by IHR, living activities such as entering, leaving and working hours can be estimated within an error range of about ± 30 minutes.

6 FUTURE WORKS

In the future, we would like to study how to control the timing of movements such as IHR swinging and nodding when talking with users and to conduct smooth conversations with people.

REFERENCES

[1] “About the present conditions and future way of watching of elderly people spreading-Mizuho Information & Research Institute (In Japanese)”, <https://www.mizuhoir.co.jp/publication/report/2017/pdf/mhir14 kourei.pdf>, (referred April.2019)

[2]“Elderly person watching business | Akiruno city (In Japanese)”, <http://www.city.akiruno.tokyo.jp/0000004255.html>, (referred April.2019).

[3] “Tapia ROBOT”, <https://mjirobotics.co.jp>, (referred April.2019).

[4] “Kibiro ROBOT (In Japanese)” , <https://www.softbankselection.jp/cart/ProductDetail.aspx?sku=4589723160100>, (referred April.2019).

[5] T. Sugihara, T. Fujinami, and K. Nakagawa, ”Effects of Cameras and Monitors on Caregivers’ Work Stress in the Group Home (In Japanese)”, The Institute of Electronics, Information and Communication Engineers, IEICE-107, No. 555, WIT 2007-100, pp.57-62, 2008.

[6] Tai Ritu, ” Let's make a natural conversation robot! Artificial intelligence made with RaspberryPi and Arduino(In Japanese)”, SYUWA SYSTEM, 2016.

[7] S. Mori, S. Yoshida, " ECHONET device adapter (Special feature / net home appliance: home network, terminal) (In Japanese)", Matsushita technical journal 47(1), pp.79-83, Feb.2001.

[8] S. Koike, “Talking Robot [Mug bot]-electronic work with Raspberry PI and Arduino-(In Japanese)”, RIC TELECOM, 2016.

[9] “ECHONET Consortium”,< http://www.echonet. gr.jp/>, (referred April.2019).

[10] “Sample to get time from NTP in JSON format(In Japanese)”, <http://etc9.hatenablog.com/entry/20110325/1301073502>, (referred April.2019).

[11] “ weather hacks(In Japanese)”, < http://weather.livedoor.com /weather_hacks/webservice>, (referred **April**.2019).

Influence of the Representation of a Conversational Agent for Supporting a Non-native Speaker to Talk with a Native Speaker

Tomoo Inoue*, and Kumiko Kawai**

*Faculty of Library, Information and Media Science, University of Tsukuba, Japan
**Graduate School of Library, Information and Media Studies, University of Tsukuba, Japan
*inoue @slis.tsukuba.ac.jp

Abstract – This paper investigates the different representations in a conversational agent for supporting a non-native speaker to talk with a native speaker. Two representations, an anthropomorphic CG agent and a voice agent were prepared for comparison. Except for the appearance, these agents were identical in terms of the function to intervene in the conversation between the native speaker and the non-native speaker and pass the next speaking turn to the non-native speaker. From the experiment comparing the CG agent condition, the voice agent condition, and face-to-face condition without an agent, initial result by a questionnaire was found that non-native speakers felt the voice agent more useful.

Keywords: Conversational agent, Second language communication, Voice agent, Non-native speaker, Turn taking.

1 INTRODUCTION

In recent years, globalization has increased the opportunities for conversation in the second language among people from different countries. Cross-cultural communication has disincentives such as differences in language and culture. Especially when a non-native speaker (hereinafter NNS) conducts conversation in a second language with a native speaker (hereinafter NS), he/she often face problems in listening comprehension of NS speech as well as in speaking, because his/her oral skills and knowledge in the second language are limited and thus difficult to make a story smoothly. As a result, there are differences in the amount of speech between NS and NNS [1], and the tendency of NS to take the initiative in conversation [2].

To improve these biases, methods and tools to support more participation of NNS in conversation have been proposed [3] [4] [5] [6]. However, most of those are not very easy to use in everyday life in terms of portability and equipment cost, as they require a PC with a display or the installation of a robot, and so on [7]. More convenient form of a support tool is desired with the expansion of opportunities for second language conversation. Audio support without video can be considered along this line. In this research, we compared the voice agent with the existing anthropomorphic agent which works in the same way.

2 RELATED WORK

2.1 Second Language Conversation Support Systems

Various studies on supporting second language conversation have been conducted recently. Majority of them are on presenting visual information such as text and images. For example, a face-to-face cross-cultural communication support system was proposed that displayed related information on the nouns appeared in a conversation [3]. A keyword sharing system to promote mutual understanding in a remote conversation between NS and NNS has been studied where keywords in the voice conversation are entered by NS for accurate and efficient content summarization and they are shown in each screen for better understanding and participation of NNS [4]. In the speech speed awareness system, the speech speed of the speaker is constantly detected and when it becomes too fast for NNS, the system notifies it through a screen [5]. Guo et al. has developed an anthropomorphic CG agent with voice to support second language conversation [6]. When a long silence occurs in the conversation between NS and NNS, the agent takes the speaking turn and talks to the NNS to pass the next speaking turn to him/her. It aims to encourage NNS utterance for better participation, which would result in better productivity in conversation.

All of these second language conversation support systems require a screen and a PC, and are not very portable. Considering that the second language conversation takes place in various places, it is desirable to develop a small portable system.

2.2 Anthropomorphic Agents and Robots for Intervention in Conversation

An anthropomorphic agent and a robot are often introduced to support conversation. Huang et al. proposed an anthropomorphic agent that participated in human conversation and presented information during the conversation [8]. From a WoZ experiment of intervention to conversation, the following four types of interventions was shown to be effective. No.1 was “provide-topic” that provides new topic to the conversation. No.2 was “more-information” that provides additional information to the topic of the conversation. No.3 was “recall-support” that supplements the information that the user has forgotten. No.4 was “discussion-support” that organizes the content of the conversation so far. In particular, the rating of “provide-

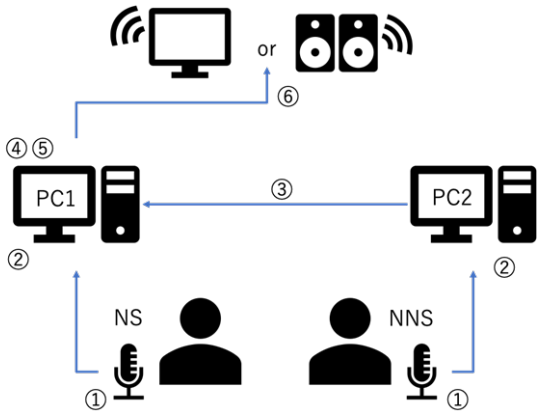


Figure 1: Configuration of the agent system

topic" was high, which indicated the demand for keeping conversation.

Akiba et al. developed a robot agent to be a facilitator in order to eliminate social imbalances in multi-party conversation [9]. The robot was life-size and intervened in the conversation between the two participants who were leading it when there was a less involved participant.

Those agents and robots also have visual appearances and need bulky equipment.

2.3 Representation of Agent and Its Psychological Influence

Research has also been conducted on the psychological effects of the form of an agent on users. Naito et al. [10] compared with a robot, a CG character, and a voice agent on the user's attitude toward information acceptance. A robot significantly increased the user's acceptance attitude, and there was no difference between a CG character and a voice agent on the information acceptance attitude. There have been other experiments to improve affinity and communication by making the agent look and act like human. According to Takeuchi et al., a human-like agent could promote relaxed dialogue with little psychological burden, but at the same time, could cause excessive expectations for its behavior [11]. Hara et al. conducted a study asking if it is possible to give more feeling of being watched by making the appearance of a camera robot eyeball-shaped. It was shown that the eyeball-shaped camera robot could make people feel more nervous and afraid than the normal camera robot. As a result [12].

Those studies indicate representation of an agent has psychological effects and influences on human behavior. In this research, we are interested in the influence of appearance of an agent that facilitate second language conversation between NS and NNS.

3 CONVERSATIONAL AGENT

The agent used in this research is a second language conversation support agent based on the turn taking rule developed by Guo et al. [6]. It is a participant in conversation. The speaking turn is passed to NNS by the agent according to the next speaker selection technique and the turn taking rule in conversation between NS and NNS.

Table 1: Agent intervention patterns

Last speaker	Intervention pattern after 2 seconds of silence
NS	Ask a specific participant for an opinion ・What do you think, NNS?
	Ask for a new opinion ・Do you have any opinion, NNS?
NNS	Ask for an utterance that supplements own opinion ・I see NNS. Could you tell me more?
	Ask for a new opinion ・Do you have any other opinion, NNS?

Figure 1 shows the system configuration. The voices of NS and NNS are by taken by microphones (1). The volume of each voice is detected by the voice detection module in PCs (2). The voice detection module notifies to the silence detection module whether received sound exceeds pre-determined threshold value (3). The silence detection module gathers information from the voice detection modules to know silence, and notifies to the Wizard, an experimenter, when silence continues for more than 2 seconds, because it is regarded as another turn even if the former speaker speaks again [13] (4). The Wizard selects an intervention speech which is most appropriate in a next turn (5). The agent start performing intervention by the selected speech (6).

As the agent's intervention, Guo et al. applied the adjacent pair pattern "Question-Response" (e.g. "What do you think, NNS?") and the additional question (e.g. "You think so, don't you, NNS?") based on the next speaker selection technique [14]. However, it might be unnatural when the agent intervenes after the NNS utterance. Therefore, in this research, the intervention patterns based on the "Question-Response" shown in Table 1 were used with reference to the intervention by Yoshino et al. [15]. In this study, we used the WoZ method to manually determine the intervention content, because it was decided based on the last speaker before the silence.

4 EXPERIMENT

We investigated the influence of the representations of our conversational agents that facilitated second language conversation between NS and NNS. Specifically, it was whether the voice agent and the anthropomorphic CG agent provided the same psychological effect on the participants.

4.1 Participants

(a) CG condition

(b) Voice condition

Figure 2: Scenes from the experiment

The participants were a total of 12 people of 6 pairs each consisting of a native Japanese speaker and a non-native Japanese speaker. Each pair has the same gender.

4.2 Design

In order to investigate the influence of agent representation, we conducted within-subject experiments under the following three conditions.

1) CG condition

The CG character agent displayed on the screen intervenes in the conversation.

2) Voice condition

The voice agent from the loudspeaker intervenes in the conversation.

3) FTF condition

Participants talk face-to-face without the agent.

Figure 2 shows scenes from the experiment. Each pair had free conversation in 5 minutes for each condition. The topic was different each time. The order of participation in the conditions was balanced by the Latin square method.

4.3 Data

The questionnaire was asked to fill out after experiencing each condition. It consisted of the items shown in from Figure 3 to Figure 6. They were the items such as naturalness of communication, ease of speech, for evaluating overall impression of communication and conversation [3] [16] [17] [18], items regarding the appropriateness of the agent intervention and whether NNS speech was promoted for the usefulness of the system, items for measuring stress in conversation, and items regarding the agent's impression for evaluating the influence of the agent representation on the participants [11] [19] [20]. Items related to an agent were put only for the CG condition and the voice condition. All items were measured in 7-point Likert scale where 1 corresponded to strongly disagree to 7 corresponded to strongly agree. It is better evaluated in the item #1 to #18 and #22 to #23 as having higher scores, while in the item #19 to #21 as having lower scores.

5 INITIAL RESULT

Figure 3 shows the result for impression of communication and conversation. Figure 4 shows the result for the usefulness of the agent. Figure 5 shows the result for the impression of the agent. Figure 6 shows the result for the stress in conversation.

The scores of NNS in the following items are lower in the CG condition than in the voice condition. They are the item #10 "I felt the agent was involved in the conversation", item #11 "The agent was useful for NNS participating in the conversation", item #12 "The agent could naturally intervene in the conversation" for the usefulness of the agent, and the item #15 "It was natural to have the agent in the conversation".

It might be possible that these lower ratings for the CG character agent compared to the voice agent reflect the over expectation for an agent with very limited capability mentioned in Section 2.3.

6 DISCUSSION

From the questionnaire survey, it is considered that the voice agent is better accepted by NNS. One of the possible reasons may be that the CG character appearance does not change and does not play any role in this conversational agent system. Only the voice element is used for intervening in the conversation. Thus the visual element is not actually useful, which could make evaluation worse than without the visual element. For the reason that the only NNS evaluated differently between the CG condition and the voice condition, it is considered that only NNS was directly interacted by the agent and thus NNS became more aware of subtle differences.

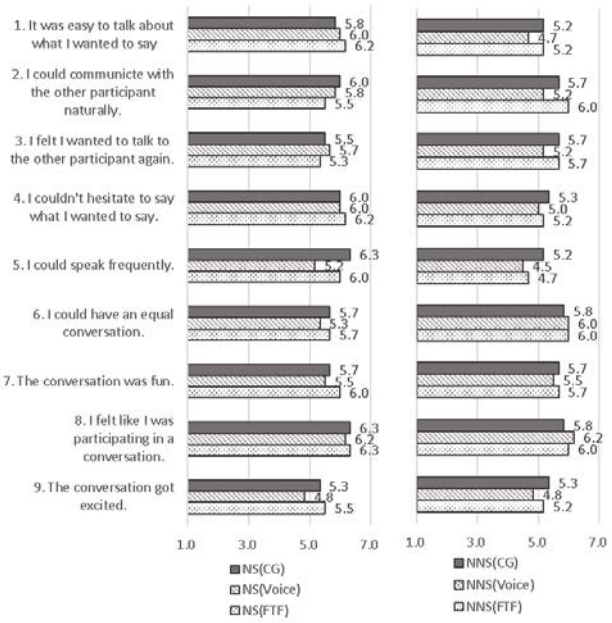


Figure 3: Questionnaire result for conversation

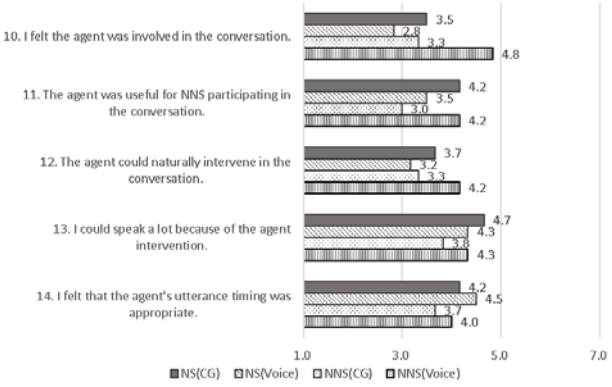


Figure 4: Questionnaire result for the agent usefulness

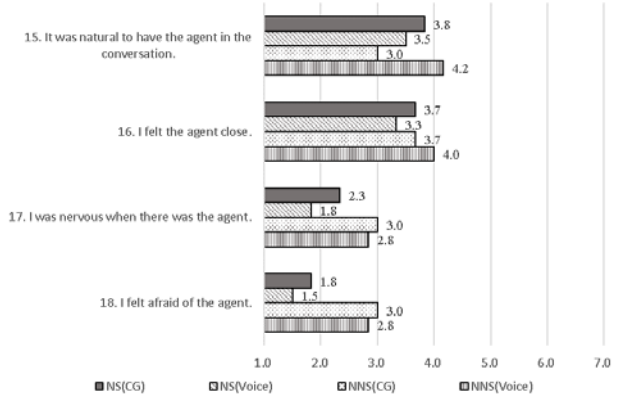


Figure 5: Questionnaire result for the agent impression

7 CONCLUSION

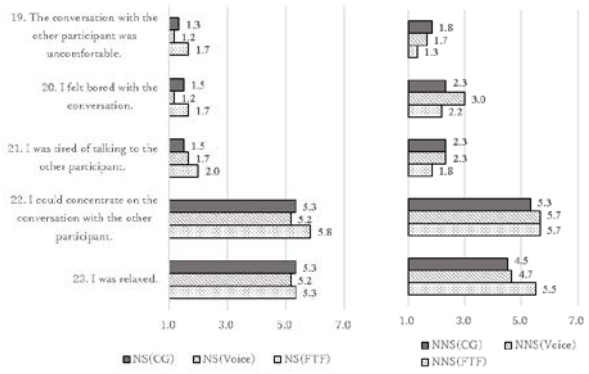


Figure 6: Questionnaire result for stress

In this research, we compare the CG character agent and the voice agent to support second language conversation. In the initial evaluation by six pairs of participants in the experiment, the intervention in conversation by the agent resulted in passing the speaking turn to NNS regardless of whether the agent was CG character or voice-only. It also made the speech balance of NS and NNS more equalized. From the results of questionnaire survey, the voice agent gave an impression to NNS that it was more useful than the CG character agent with the similar functions. In the future, we plan to increase the number of participants in the experiment to increase the reliability of the results.

REFERENCES

[1] W. Zhu, Interaction and Feedback in Mixed Peer Response Groups, Journal of Second Language Writing, Vol.10, No.4, pp.251-276 (2001).

[2] T. Hifumi, Relationship of Linguistic Processing and Psychological Processing in Conversations Between Native Speakers of Japanese and Nonnative Speakers, Japanese Journal of Educational Psychology, Vol.47, pp.490-500 (1999).

[3] K. Okamoto, and T. Yoshino, Development of Face-to-face Intercultural Communication Support System using Visualized Keyword of Conversation, Proceedings of Forum on Information Technology, Vol.8, No.3, pp.393-396 (2009).

[4] H. Hanawa, X. Song, and T. Inoue, Communication Support by a Native Speaker's Entering Keywords on Discussion with a Non-native Speaker -- Effects of Entering Keywords by a Native Speaker on Audio Conference --, IEICE Technical Report, Vol.116, No.31, pp.139-144 (2016).

[5] J. Ye, and T. Inoue, A Speech Speed Awareness System for Non-Native Speakers, Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion, pp.49-52 (2016).

[6] Z. Guo, M. Tang, and T. Inoue, A Conversational Agent for a Non-native Speaker to Talk with a Native Speaker, IEICE Technical Report, Vol.117, No.509, pp.107-112 (2018).

[7] S. Yamada, Are Robots Necessary for Human-Agent Interaction?, Proceedings of the 23rd Annual Conference of the Japanese Society for Artificial Intelligence, Vol.23, No.1G2-OS3-11 (2009).
[8] H. Huang, S. Otogi, R. Horita, and K. Kawagoe, Toward a Guide Agent Who Actively Provides Information in Multi-user Conversation – A Trial in Finding the Timings for Intervention, Transactions of Japanese Society for Artificial Intelligence, Vol.31, No.1, 10.1527/tjsai.DSF-514, 2016.
[9] I. Akiba, Y. Matsuyama, and T. Kobayashi, Procedures of Obtaining Initiatives for Multiparty Conversation Facilitation Robots, IPSJ SIG Technical Reports, Vol.2013-SLP-97, No.10 (2013).
[10] H. Naito, S. Kawamura, and Y. Takeuchi, An Empirical Study of Indicating Reliable Messages in Real World Interaction, The Transactions of the IEICE. A, Vol.92, No.11, pp. 840-851 (2009).
[11] Y. Takeuchi, and Y. Katagiri, Inducing Social Agreement Responses Toward Interface Agent, IPSJ Journal, Vol.41, No.5, pp.1257-1266 (2000).
[12] T. Hara, and K. Terada, Gankyu robot ni yoru hishikan ga doutokusei koujou ni ataeru eikyo, IPSJ SIG Technical Reports, Vol.2015-HCI-164, No.1 (2015).
[13] J. Rubin, A Review of Second Language Listening Comprehension Research, The Modern Language Journal, Vol.78, No.2, pp.199-221 (1994).
[14] H. Sacks, E. A. Schegloff, and G. Jefferson, A Simplest Systematics for the Organization of Turn-Taking for Conversation, Vol.50, No.4, Part1, pp.696-735 (1974).
[15] T. Yoshino, M. Yatsushiro, Y. Takase, and Y. Nakano, Intervening in Multiparty Conversations by Conversational Agents based on Dominance Estimation, Proceedings of the 29th Annual Conference of the Japanese Society for Artificial Intelligence, Vol.29, No.1I4-4, pp.1-3 (2015).
[16] R. Liu, and T. Inoue, Application of an Anthropomorphic Dining Agent to Idea Generation, Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, pp.607-612, (2014).
[17] T. Inoue, Naturalistic Control of Conversation by Meal: Induction of Attentive Listening Attitude through Uneven Meal Distribution in Co-dining, Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, pp.601-606 (2014).
[18] T. Takahashi, M. Hamasaki, and H. Takeda, Web Community System Mediated by Avatar-like Agents, Proceedings of the 15th Annual Conference of the Japanese Society for Artificial Intelligence, 1F1-10 (2001).
[19] Y. Hayashi, and K. Miwa, Cognitive and Emotional Characteristics in Human-Human / Human-Agent Interaction, The Transactions of Human Interface Society, Vol.10, No.4, pp.445-455 (2008).
[20] Y. Hayashi, E. Cooper, V. Kryssanov, A. Urao, and H. Ogawa, Psychological Characteristics on Communication with a Conversational Agent - A Study on Schema and Embodiment -, Transactions of Japan Society of Kansei Engineering, Vol.11, No.3, pp.459-467 (2012).
[22] A. Shamekhi, Q. V. Liao, D. Wang, R. K. E. Bellamy, and T. Erickson, Face Value? Exploring the Effects of Embodiment for a Group Facilitation Agent, Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp.391-403 (2018).
[23] S. K. Maynard, Discourse Analysis, Kuroshio Shuppan (1993).
[24] N. Yoshida, K. Takanashi, and Y. Den, Taiwa ni okeru aiduchi hyougen no nintei to sono mondaiten ni tsuite, Proceedings of the Annual Meeting of the Association for Natural Language Processing, Vol.15, pp.430-433 (2009).

Session 7:
Industrial Applications
(Chair: Hiroshi Inamura)

Project Manager Development Methodology through PMO

Akihiro HAYASHI*

*Department of Information Design, Shizuoka Institute of Science and Technology, Japan
pixysbrain@gmail.com

Abstract Using Capability Maturity Model Integration (CMMI), we are working towards achieving maturity level 3. However, if the activity breaks down, the resources of the indirect department become unnecessary. In order to hold their role as CMMI activities, it often remains a project management office (PMO). However, the PMO's role has not been specifically defined.
In this case, after pointing out the challenges, we propose the role of process improvement and the method of transferring technology to the project manager as an in-house consultant for PMO. When this method is applied to actual development sites that completed CMMI level-3 activities, it is expected to effectively improve the reuse rate and implementation of autonomous improvement of Project Manager (PM).
Keywords: CMMI, Project Management Office, Knowledge Transfer

1 INTRODUCTION

According to the Capability Maturity Model Integration (CMMI) Maturity Profile [1] published by the CMMI Institute, CMMI [5] appraisals have been conducted continuously in many countries around the world every year. CMMI is supported worldwide as the most reliable tool in the field of process improvement.
When CMMI was introduced in Japan around the year 2000, the level of competition was that of a temporary boom and focused on not only improving process but also achieving CMMI Maturity level 3. However, it has now been recognized as a perpetual and continuous activity as part of the organizational culture.
When the organization aims to achieve CMMI Level 3, the organization needs a lot of resources to conduct process improvement activities. This resource is named the Engineering Process Group (EPG).
However, once CMMI Level 3 is achieved, the next goal is to maintain this level of surveillance for implementation once every three years. The activity itself is not profitable: From an organizational point of view, EPG is an indirect operation and a cost center. After achieving the preliminary organizational goals, personnel in the indirect department should be transferred to a profit center or assigned a new meaningful role as an indirect department. However, CMMI activities need to be continued and the EPG cannot be abolished.
Therefore, the issue is to define the role and responsibilities of the EPG in the second stage of process improvement. In this context, the study proposes the way in which the project management office (PMO) may be established after achieving a CMMI maturity level and evaluates its validity through application to the actual organization.
As part of a previous study in this field, the authors proposed establishing a methodology for process improvement activities across the entire organization that will continue after achieving CMMI level 3. However, we have cautioned that CMMI's appraisals have rated the achievement of process improvement across the organization based on an evaluation of a few representative projects. We have proposed a methodology for disseminating CMMI activities throughout the organization [2]. However, at this stage, it has not been proposed as a PMO-based activity.
Various consulting firms have proposed that the organization of the EPG be transferred to the PMO after the introduction of CMMI activities. These proposals emphasize the cost improvement through in-house consulting by PMO rather than paying consulting fees to external consulting firms [3], [4].
However, there is no case where experienced PMOs give technology transfer to PMs and propose continuous process improvement using Project Based Learning (PBL).

2 ROLE AND ISSUES OF PMO

After achieving the organizational goals of CMMI activities, the EPG often turns to the PMO to continue to support project management and process improvement. The PMO is defined as a team that supports individual project management in an organization across the board. In general, the role of the PMO is as follows:

- Standardization of the project management system
- Human resource development including training on project management
- Project management support
- Coordination of resources and costs between projects
- Development of a project environment tailored to individual companies
- Other related project management tasks

Because the PMO is an indirect operation, appropriate rationale for the survival of the PMO is required. The long-term role may be CMMI surveillance measures, and the short-term role, support activities such as reducing overload of PMs. If the PMO absorbs the work associated with CMMI activities, it can survive by providing a value-added role.
However, the PMO often does not successfully reduce the burden of the project manager (PM). Here, we will discuss the role of the PMO through a case when a company that achieved CMMI Level 3 established a PMO.

2.1 PMO Failure Case

Company Z started activities of ISO9001 and CMMI about 10 years ago, passed ISO9001 certification registration 3 years later, and achieved CMMI maturity level 3 in 8 years.

The management has a strong commitment to CMMI and about 10 EPGs were in operation. Company Z's EPGs were focused on project guidance and included some fixed-term employees who had passed the retirement age of 60.
In the project construction period, development man-hours are set in line with the estimate. About 15% of the entire project term is appropriated for management operations [6], [7]. Furthermore, introducing CMMI increases overhead man-hours by about 10%. Therefore, when CMMI is introduced, about 25% of the total project man-hours would go toward the management, and the delivery date may be delayed due to unnecessary management costs. Therefore, the reduction of management man-hours after achieving CMMI level 3 has become an issue.
The PMO replaced management-related activities with project support. If the delay in delivery date is resolved by this PMO, the PMO activities are considered as value-added activities and it would be a position with merit for both.
PMO was substituted for the following management work as a specific effort to reduce the load on the PM:

- Create weekly reports
- Create progress meeting minutes
- Create the trail necessary for the appraisal

2.1.1. PMO Reporting

The project situation is reported directly to the Executive Committee at a meeting held every Monday. If the project status is not reported in a timely manner by all executives attending the meeting, the management cannot manage the project status.
Until now, the PM has delivered project reports in Company Z. However, in small-scale organizations, the creation of weekly reports tends to be delayed because the PM is involved in both management and development. There have been many cases where reports were lacking in politeness. Since the PMO has EPG knowledge, they have learned to prepare the risk and problem management sheets. They have delivered plausible reports and there have been no delays in reporting. They are also valued by the Executive Committee, and the PMO was recognized as a value-added group.

2.1.2. Create Progress Meeting Minutes

Since the management meeting is held on Mondays, the project progress meetings are held on Thursdays. The project progress meeting reports regular issues and issues that occurred in the week. The PMO attends the progress meeting of the project-in-charge and provides minutes of the meeting. Although the progress meeting had been held until then, creating the record was the individual responsibility of participants, and no official minutes were captured for the project.
Company Z was a small-scale organization, and was not able to prepare, review, and approve official minutes in weekly progress meetings for small-scale projects.
However, CMMI has processes that require the creation of minutes, such as project progress management, official reviews, and peer reviews, and they tend to be short. The PM therefore took charge of this meaningful activity, but he was not interested in it.

2.1.3. Creation of the Evidence for Appraisal

To pass three-year surveillance, evidence of ongoing practice must be presented. For example, to pass the configuration management process, a configuration management plan must be prepared, and the evidence for carrying out the activities according to the plan must be presented. In addition, when modifying the source code, there is a procedure to check out the source code from the repository, modify the file, hold the configuration management committee, analyze the impact, and check in.
Configuration management was performed at Company Z, but there was no custom of creating forms and recording the contents of changes each time. The PMO substitutes for the difficult task of countermeasures and highlights the added value of the PMO.
In small organizations, the PM is relatively young and the PMO was able to compensate for their lack of skills. A proper weekly report has been submitted to the Executive Committee every week. Surveillance that will be conducted once every three years will also pass with confidence. Since the PM burden has been reduced by PMO activities, this method seems to have benefits for both parties.
However, in this method, the PMO is only a substitute for the PM. Furthermore, the PMO sometimes reported the causes of problems and measures to prevent recurrence, to the Executive Committee, without the consent of the PM. The management said, “Every time a problem occurs, we ask for measures to prevent a recurrence, but it is not improved at all.” Measures to prevent recurrence were written by the PMO without confirming the problem with the PM.
The lesson learned from this case is that although the involving the PMO is a value-added activity, the method should be aligned to the process of reducing the burden of the PM. The role of the PMO is to support process improvement and not of a PM substitute.

2.2 Issues to be Solved

The importance of process improvement and project management is recognized widely. Even if one hires an external consulting company and develops an EPG in-house, no methodology has been proposed for how to use the resources after the activity has ended. Defining the PMO’s value-added activities remains an issue to be resolved.

3 TRANSFER METHOD OF PROJECT MANAGEMENT TECHNOLOGY BY PMO

In this research, after defining the role and responsibility of the PMO, we propose a method of continuous improvement of the process by PMO.

3.1 Basic policy

In this research, the PMO not only reduces the burden of the PM but also inherits the EPG and is responsible for post-approval in-company process improvement. Unlike the EPG, which is a permanent organization, the PMO has the role of transferring the completed process to the next-generation PM.

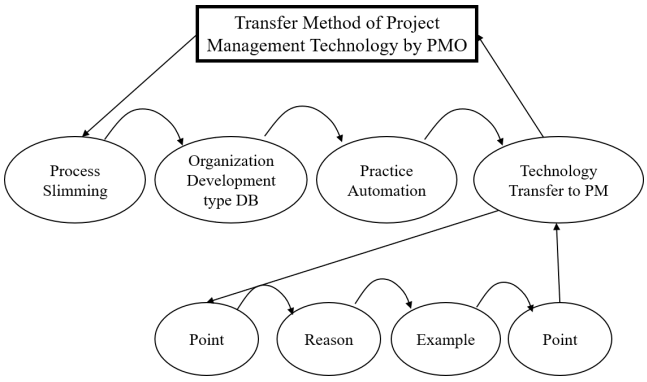


Fig. 1 Framework for the PMO work in this study

3.2 Framework

In this study, we propose a framework for the PMO work that consists of process slimming, organization development-type database creation, practice automation, and technology transfer to PM. The first three are organizational-level activities, and the fourth, a project-level activity. Technology transfer to the PM follows the PREP method. The PREP method first conveys the conclusion, then explains the reason, reinforces the reason with the case, and finally presents the conclusion again. This is then classified based on importance of activities (Point), reasons (Reason), specific examples (Example), and situations where activities are important (Point). This framework is shown in Figure 1.

3.2.1. Process Slimming

Achieving CMMI's maturity level cannot be avoided through CMMI's "embedding." However, CMMI has been developed as a procurement model for the US Department of Defense, so a large-scale project is assumed. After achieving the CMMI level, the company selects only what is needed and slims the process by removing the excess. However, there are many practices that are not as familiar as domestic practices.

3.2.2. Organization Development type Database Creation

In a system development organization, projects with the same degree of development and difficulty are repeatedly implemented for similar customers. By classifying the development type of the organization and converting it into a database, standardization for each customer is possible.
Next, the individual information of the classified forms is sanitized, the past information is accumulated in a database for each customer, and the reuse environment is constructed.

3.2.3. Practice Automation

When creating a form for each process area, the form is retrieved from the organization's database and the organization management profile is automatically embedded. The basis of project management is PDCA (Plan, Do, Check, Act). At the project start stage, make a project plan document and execute the project according to the plan. However, in the upstream process, the load on the PM is large, and there is no time to create a detailed project plan. As a result, project management cannot be planned. Database forms classified by organization type can be used without much change in repeatable projects.
Therefore, we automatically generate forms as much as possible using historical project information accumulated in the organization.

3.2.4. Technology transfer to PM following the PREP method

Technology transfer is to transfer management skills to inexperienced PMs through development projects. Generally, the practice is to fill in the form given, even if it was to create a plan for an unknown project. This practice does not outline the intention clearly.
In this study, we use the concept of the PREP method to make the PM learn the management by the PDCA cycle. The PREP method first conveys the conclusion, then explains the reason, reinforces the reason with the case, and finally presents the conclusion again.
First, the purpose of the practice is clarified, and then its necessity is explained through concrete examples. By explaining the purpose of implementing the practice once again, and being convinced by the PM, management skills are passed on.

4 EVALUATION OF APPLICATION

In this chapter, we verify the effectiveness of our proposal through a case of actual application of this framework.

4.1 Case

Company A has approximately 400 employees and is developing control systems for the automotive industry. About half of the employees are in charge of development. The project period is about three months, and many are small projects with less than five project members. One engineer can participate in projects up to four times a year. Usually, about 10 projects of the organization are in operation.
Company A hired a foreign consulting firm to introduce CMMI. In order to implement unified practices in the organization, CMMI is required to create organizational standards, and each project should carry out project management according to the standards.
Company A obliges to execute the project by using CMMI's development work standard and Excel-based format, brought in by the consulting company.
Development work standard corresponds to the process area of CMMI, and refers to "requirement definition standard," "project plan standard," "progress control standard," etc. An Excel-based format refers to work products such as "project plan document" and "Work Breaking Structure (WBS)" created when implementing project management using development standards.
However, this consulting company created consulting materials based on the project management standard originally used by the US headquarters. It is a development standard for large-scale projects. It is hard to say that it is suitable for small-scale companies like Company A.

Company A compared with other companies' proposals at the selection stage of the consulting company. Only this company provided a set of development standards and formats for use in CMMI. Company A adopted this consulting company because there was no effort to create such development standards and formats from scratch.
Company A achieved CMMI Level 3 and is preparing for surveillance after 3 years. Ten employees who worked as EPGs until CMMI Level 3 were considered. Four were transferred to the department, and the remaining six became PMOs who were positioned in the project support team.

4.2 Application of this framework

4.2.1. Process Slimming

Company A used development standards and formats provided by a consulting company, until CMMI Level 3 was achieved. The CMMI Achievement Appeal is conducted by interviews and document review. The questionnaire is published in advance. If all the forms given by the consulting company were prepared and the questions were answered, CMMI Level 3 could nearly be achieved.
In the first round of CMMI activities, Company A prepared the form of the consulting company on a persuasive basis in preparation for the appraisal. However, after the official appraisal, it was found that several forms were created just for the preparation of the appraisal, and nearly 20 were eliminated.

Therefore, the process was streamlined by eliminating forms that were considered unnecessary. Table 1 presents an example.

4.2.2. Organization Development type Database Creation

Many customers of Company A are major Japanese automobile manufacturers. Customers are divided into company T, company H, and company N. For customers, projects of the same size and the same degree of difficulty are implemented in a repeating manner. Therefore, the development type was classified for each customer to be delivered and put into a database.
After a certain period of time, the most commonly used format among the three customers was taken out, and customer names and personal information were sanitized, standardized, and stored in the database.

4.2.3. Practice Automation

Company A is doing SAP's ERP (Enterprise Resource Planning), the year after CMMI's level 3 activity has ended. Profile information of the project of the organization is entered into ERP when the contract is established. The ERP information is then downloaded to an MS-Excel file.
When a development project creates a project plan, ERP information is acquired by MS-Excel, and the information is automatically reflected in the standardized project plan.
Next, similar databases were searched based on customer, project schedule, difficulty level, etc. which were classified by T, H, N companies, and those that could be reused at the project planning stage, were also extracted; for example, WBS, risk management ledger, configuration management ledger, etc. If these are similar projects, the same task occurs in almost the same process. As a result, the resulting risk management ledger could be reused.

4.2.4. Technology transfer to PM following the PREP method

At Company A, when a project was launched, one PMO was assigned to the project and played the role of a tutor for PM, transferring the technology to the PM by the PREP method.

(1) Project report (Point)

The role of the PM is project management. However, if young people become PMs without management experience, they would not know what must be done in every phase of the project life cycle. Therefore, the PM was made to participate in the actual project, and the importance of the project management method was highlighted through the PBL method.
Project management is performed in the PDCA cycle. Table 1 shows the relationship between PM and project members throughout the project life cycle.

The project planning and progress management phases are important management skills of the PM. Therefore, Company A taught the progress meeting in the first project to the PM, and in the second cycle, the PM learned to make a project plan. In the third cycle, the PMO became an observer and decided to implement project management through the PM.

Table 1 PDCA Cycle of Project Management

	P	D	C	A
PM	Create a Project Plan		Progress Meeting	
Member		Development		Correction

(2) Gathering information for reporting (Reason)

PM is an intermediate manager. The middle manager refers to the position that reports to the upper management of the organization. In order to report to upper management, the PM correctly understands the current situation of the project and reports the problems in a timely manner.
At Company A, when the project was successful, weekly reports were made by briefly explaining the issues described in the issue management sheet. However, if there is a delay in progress or a technical bottleneck, the upper management will request a detailed report. For this purpose, PMs need to hold weekly progress meetings and collect the information requested by upper management.
In the first cycle, the PMO chaired the progress meeting and held the progress meeting using the CMMI's agenda. Company A is supposed to update WBS before coming to the weekly progress meeting. At the progress meeting, we reviewed the Earned Value Management (EVM) value of the project and interviewed each project member's progress, problems, issues, and risks. This information was collected and reported to the reporting manager.

In the second cycle, I participated in the PMO from the preparation of the project planning documents. Project management is performed by the PDCA, so project plans are prepared first and then managed for progress. However, even if you do not know the purpose, this cannot be understood through merely creating the form. Therefore, the PM had experience in managing the progress meetings and understanding the type of management to be followed with different projects. The project plan was prepared after the meeting.
Company A defines the life cycle process at the project planning stage. Tasks performed in every process are described in the WBS. The output for each process is described in the configuration management register. The first activity is to confirm the output of the process at the actual progress meeting and the second activity is to plan the output of the process at the project planning stage.

(3) Form to be created (Example)

The practices to be implemented as project management are clear. The PM needs to learn all the practices, but as a concrete example of technology transfer, he followed the practice that reversed the PDCA cycle as mentioned above.

- Report content requested by upper management
- Gathering information for accurate and timely reporting
- To manage progress meetings and to collect information weekly
- Enter Gantt chart, WBS, issue management ledger, risk management ledger interviewed at a progress meeting
- List of work products to be created in each process
- Process plan such as configuration management ledger that defined the output for each process
- A project plan document summarizing each process plan

By making a project plan document at the end of the technology transfer, we have made it possible to understand different process for the entire project life cycle.

(4) Report for clarity (Point)

For organizations achieving CMMI Level 3, the Project Progress Meeting agenda is in the organizational standard. At the progress meeting, interviews with project members will be conducted according to the agenda, and the minutes of the progress meeting will be recorded. A project management trail is kept and used as evidence when conducting an appraisal.
However, the practices required by CMMI are minimal. Even if the information is collected by holding a progress meeting in line with the agenda, it may not be consistent with the reports that the upper management wants to hear. There are times when you call the site many times during closed hours.
The project report is finally reviewed at the management meeting. Even if the project was managed according to CMMI, number of issues were pointed out at the management meeting, and new issues were generated. As a result, management also found that CMMI was not useful.
However, the management complained that extra work would be required in the following process: project plan document → progress meeting → reporting to upper management and implementation. We set the appropriate agenda at the progress meeting and learned how to collect the necessary information.
It was understood that the project plan document would not be prepared by management if it was inaccurate, in order to manage the reality at the progress meeting. The project was planned on more accurate grounds.

4.3 Application results

4.3.1. PM recognized autonomous management

The PMO is a role set up for a fixed period, and not permanently. The number of indirect personnel should be reduced if the PM manages the project autonomously. This activity was started, and technology transfer activities were conducted for the PM. After three rounds of short-term projects were evaluated using the PREP method, the PM, who could perform autonomous project management by himself, finished with the PMO's support. This situation is shown in Figure 2.

4.3.2. Format reuse rate

At first, the project plan was prepared from the beginning using the format which closely followed the given blank sheet. In the third year of introducing this activity, about 60% of the preparation of the project plan had been automated. It became better for the PM to manually input the remaining 40%. This is because project planning documents and other forms used for each client is collected, converted to DB, sanitized, and reused.
Necessary information for clients is almost the same, so it can be reused. Information required for each project automatically inputs from the Excel file output from ERP. Nearly 48% of these improvements have been automated. This situation is shown in Figure 3.

4.3.3. The contradiction between measures against appraisal and slimming

Company A's form was brought in by an external consulting firm for appraisal. At Company A, the form was streamlined and made easy to read.
CMMI passes if a practice called REQUIRED is implemented. There is no problem in slimming the “nice to have” part. This caused the metamorphosis. It is inevitable that this activity “fits in the mold.” It was decided that the unnecessary aspects of the organization would not be included in the form.
We repeated the procedure of slimming down a part that was not actually used even though it was introduced as a full set. Moreover, only the passing of the appraisal was maintained. As a result, a lean development standard was completed, conforming to the spirit of CMMI.

4.3.4. Inconsistency of form creation automation and PDCA

Project management is to create a project plan document and manage it according to the plan. If more than 60% of the project plan document is generated automatically by automating the form, the perspective of planning the project

is missing. There is a meaning in labor saving. However, if you do not plan the project carefully in the upstream process of the project, it will be overturned.
About 60% of the format automatically generated in this study was client information or the project profile output from ERP. The individual information on the project is about 40% of the rest. This part was created manually by the PM.
There is an idea called Pareto's law. If 40% of the project plan is hand-crafted by the PM, it will be sufficient as a project plan. Specifically, information on project schedule and process, allocation of resources and assignment of personnel have completed.
The project plan of Company A is created with an MS-Excel of about 12 sheets in one book. In terms of the number of sheets, eight sheets were automated. The remaining 4 sheets were created by the PM. This part implements the project plan firmly and is in line with the spirit of PDCA.

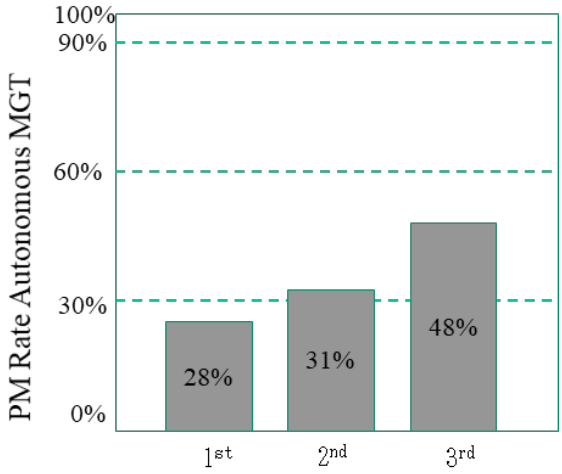


Fig.2 PM Rate recognized autonomous MGT

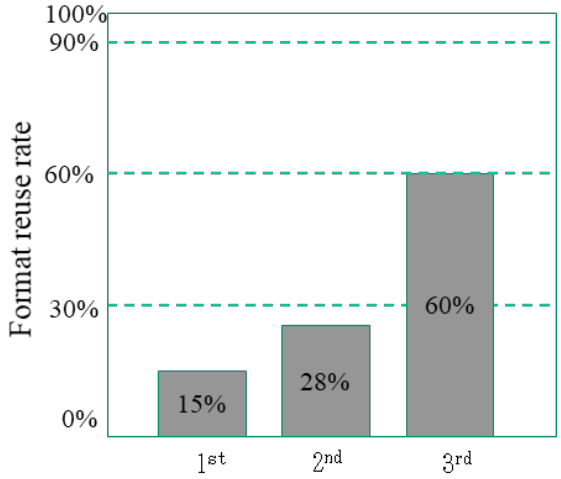


Fig.3 Format Reuse Rate

5 CONCLUSION

This study proposes that, after the CMMI's activity is completed, the resources that were active in an EPG continue to function in the PMO.
As it is difficult to achieve CMMI Level 3 alone, we hire an outside consulting company. However, it is not desirable to keep paying high external consulting fees. It is therefore desirable for knowledge-rich EPGs to be in charge of process improvement as in-house consultants and to transfer technology to young PMs.
Many organizations set up PMOs after CMMI activities. However, they prepare weekly reports and minutes and collect CMMI appraisal trails, which does not fulfill the roles and responsibilities of effective in-house consulting.
In this study, we focused on the role of process improvement in PMO. The activities required for the organization were identified and narrowed down. Next, by classifying the forms for each customer, we added it to the database and made it reusable. The knowledge was transferred to young PMs using the PREP method in which the PMO played a significant role.
In a company that adopted the proposed method, the process reuse rate increased, and the PM began to carry out project management after acquiring knowledge of process improvement, which was considered to be an effective method.

REFERENCES

[1] CMMI Institute, Process Maturity Profile: CMMI Institute, 2013.
[2] Akihiro HAYASHI, Nobuhiro Kataoka, “A method to Stabilize Process Improvement Activities—Proceeding to Conduct Process Improvement Activities after Achieving CMMI Level 3—”, IEICE, D, vol. J98-D, no. 9, pp. 1237-1246. 2015.
[3] Koji Watanabe, Process Improvement to lead Project Success, CSK, 2005.10.13
[4] Hitachi Solutions, Software Quality Assurance, 2011.9.9
[5] CMMI Product Team, CMMI for Development, ver.1.3, Improving processes for developing better products and services: CMU/SEI-2010-TR-033. Software Engineering Institute, 2010.
[6] Shoji Okamura, Zettaini Chienshinai Project Sinchoku kanri, Scientific Management by Master Schedule, Earned Value Management, PMS, Nikkei BP, 2010.
[7] Ten Step Japan, Project Management Process (Ten step process, http://www.tenstep.jp/cms/

Consecutive Position Control for Spindles in CNC Equipment

Kazuhiko Tsutsui[*†], Jun Sawaki[*], Naoya Chujo[††] and Tadanori Mizuno[††]

[*]Mitsubishi Electric Corporation
[†]Graduate School, Aichi Institute of technology, Japan
[††] Aichi Institute of technology, Japan

Abstract - Recently more advanced functions have become necessary for machine tools in order to achieve high productivity in the field of factory automation.

In particular, performance and functionality of spindle motor have been becoming more important.

In the past, only speed control was necessary for spindles. However, switching between position control and speed control has become necessary to allow for various machining modes for the increasing number of systems with advanced spindle functions. The time and complexity required for switching control loops are bottlenecks for improvement in productivity of machine tools.

In this research, we propose an approach to perform consecutive position control for spindles to resolve these bottlenecks.

Keywords: Spindle motor control, Multi-tasking machine, Consecutive position control, Tracking delay compensation

1 INTRODUCTION

CNC equipment consist of the CNC controller, actuators, amplifiers, and detectors. To run the user's machining program, the CNC controller generates position/speed commands sent to the feed axes and spindles (tools) that constitute the machine coordinate system. Actuators such as servo motors and spindle motors are used to drive the feed axes and spindles. Servo amplifiers and spindle amplifiers are used to supply variable power to the actuators. Besides, servo amplifiers and spindle amplifiers are used to control the speed, position and output force of the actuators. Detectors such as encoders and linear scales are used to feed back the position/speed of operating parts of machines or motors.

The CNC controller sequentially analyzes a G-code program that include description of the machining path, feed speed, and number of rotations of the tool, and generates commands of the travel distance per unit time for each feed axis. The CNC controller also generates speed commands for spindle motors according to the cutting conditions.

While servo/spindle amplifiers are connected one-on-one to motors with power lines, serial communication through one network interface is used between the CNC controller and multiple amplifiers.

Recently, the spindle motor control has been more complex to achieve high productivity. The servo axes don't have to change the control mode. It means that the servo axes are controlled by position loop consecutively. On the other hand, spindle motor control is required to use properly the speed loop control and position loop control. Although spindle amplifier is only needed to control the spindle motor speed and output power on the process of turning and milling, it is needed to control the angle position of cutting tool and to synchronize with other servo axes on the process of tapping C-axis mode for multi-tasking machine.

In the former systems, the mechanical angle of spindles had to be adjusted every time when the machining mode was changed. As it was necessary to recalculate a designated position, it was not possible to reduce the machining time.

Hence, we propose an approach to perform consecutive position control in spindle amplifiers for spindle motors regardless of the machining mode.

This approach reduces extra sequential control to change the machining mode, ensures robustness against disturbances, and enables establishment of the spindle control system for high-accuracy machining.

2 RELATED TECHNOLOGIES

2.1 Basic Configuration of CNC Equipment

Multi-tasking machines such as the one shown in Figure 1 are increasing in recent years.

Each axes are driven by the servo motors connected to the ball screw in the machine. The tool used for cutting is attached to the spindle head and driven by the spindle motor. The CNC controller and the amplifiers which is used to control the speed and position in addition to power supply for the servo and the spindle motors are all installed in an electrical enclosure.

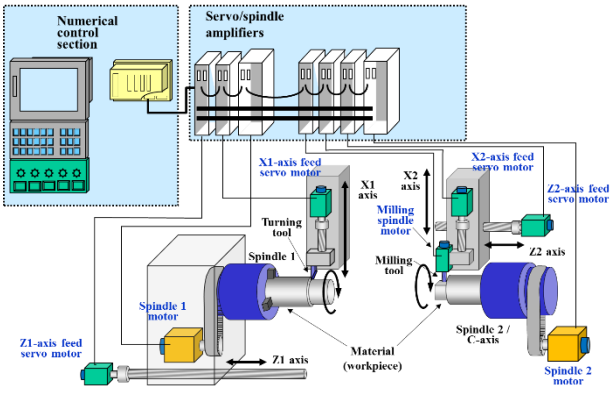


Figure 1: System configuration example of multi-tasking machine

The path of the cutting tool affects accuracy of the workpiece directly. Therefore, it is significant for servo motors to ensure robust command tracking for position and

feed speed control commands with minimal errors by suppressing load disturbances such as cutting reaction force or friction in machines. It is also important to keep the synchronous accuracy and the same response level for X, Y, and Z axes. Otherwise, the tool may follow the wrong path not intended by the CNC controller, resulting in unsatisfactory machining accuracy.

2.2 Basic Servo/Spindle Control Architecture

We adopt the architecture shown in Figure 2 for servo and spindle control loops. The position feedback and speed loops are implemented inside the amplifiers (distributed control) and the network communication that involves a substantial amount of dead time is established outside of the control loops.

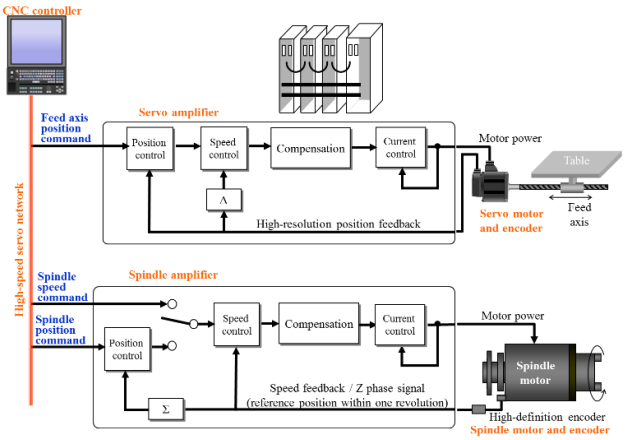


Figure 2: Servo/spindle control loop.

2.3 Issues Identified with Spindle Control

Multi-tasking machines that perform multiple operations prevail in recent years. Position control and speed control for the spindle motors are switched in multi-tasking machines according to the machining program (refer to the spindle amplifier section in Figure 2). In addition to operations such as turning, milling, or drilling by spindle rotation control, it is necessary for multi-tasking machines to perform operations such as C-axis control or synchronous tapping (C-axis control: synchronous position control with the servo axes).

2.3.1 Time required for switching the control methods

When the spindle requires only speed control in the machining mode such as turning, neither the CNC controller nor the spindle amplifier handles position data such as tool tip angles.

In this case, to change the control mode from turning to milling (C-axis control), it is necessary to stop the spindle once, perform home position return, and detect the tool tip angle as shown in Figure 3.

Another example is shown in Figure 4. To change the machining mode from turning one workpiece chucked by two spindles to milling (C-axis control), home position return is required for both spindles. The chuck of each spindle is opened to release the workpiece and home position return is performed one after another for two spindles. Then, the chuck is closed and the workpiece is held by the spindles again, the C-axis angle is calculated, and milling is started. It takes several to tens of seconds each time to switch between two modes. Therefore, repeating switching has significant impact on productivity.

Moreover, the switching process sequence control involves complex programming.

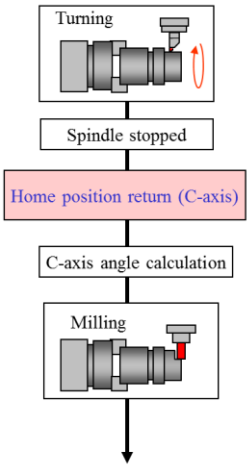


Figure 3: Switching from turning to milling (C-axis).

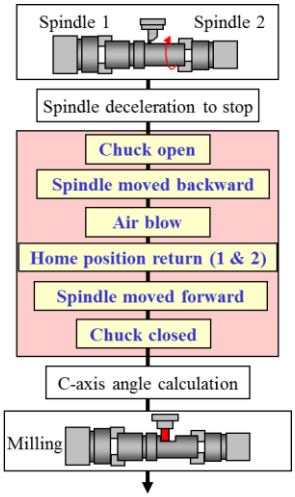


Figure 4: Switching from spindle-linked turning to C-axis-linked milling

2.3.2 Consecutive position control in spindle amplifier

The following section describes the issues involved with consecutive position control in spindle amplifiers in the same way as in servo amplifiers.

The tool attached to the spindle is frequently changed according to the machining conditions. The spindle motor must be decelerated and stopped once every time the tool is changed, and accelerated back to the maximum speed after the tool change. The number of spindle rotations per minute ranges from several thousands to tens of thousands and it takes several to tens of seconds each time to accelerate or decelerate the spindle. One of the requirements demanded for spindle motor control is to reduce the acceleration and deceleration time as much as possible.

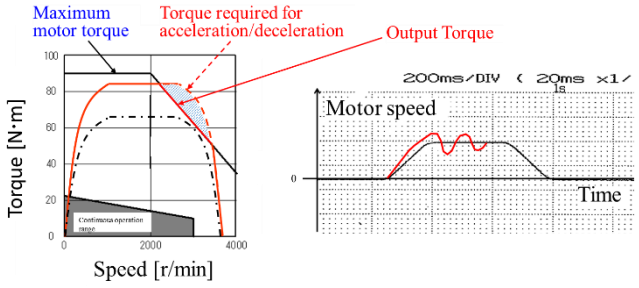


Figure 5: Abnormal acceleration/deceleration waveform of position loop control.

Figure 5 is a diagram to show the maximum motor torque. The horizontal axis of the graph represents the motor speed.

In general, the maximum torque generated by motors is reduced due to saturation of the induced voltage in the high-speed region even though the maximum current value of the amplifier is not changed.

Under these circumstances, if a program is built with a time constant that would require the torque exceeding the maximum motor torque as shown in Figure 5, position control to follow the command is not achieved. Then, position errors are accumulated to cause overshooting, resulting in vibration or mechanical impact.

On the other hand, acceleration to the maximum speed or deceleration to a stop in a shortest possible time is desirable in the machining mode without position control as mentioned above. Therefore, the utmost motor torque must be applied for acceleration/deceleration. As shown by the solid line in Figure 6, in the case of speed loop control, the utmost motor torque can be applied for acceleration/deceleration without overshooting since position deviations are not accumulated even though a program is built with a time constant that would allow the torque require for acceleration/deceleration to exceed the maximum motor torque.

However, it is not possible to control the cutting tool's angle position on the spindle since position control is not performed in this case. For that reason, position control and speed control are switched according to the machining mode for spindle control in preceding systems.

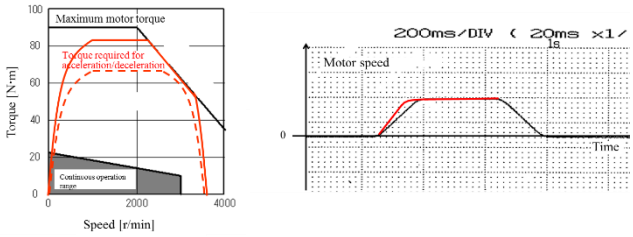


Figure 6: Acceleration/deceleration waveform of speed loop control.

3 PROPOSED METHOD

3.1 Outline of Tracking Delay Compensation Control

As mentioned above, a major issue for consecutive position control is overshooting. Overshooting occurs when position control to follow the command is not achieved, resulting in position deviation, during acceleration/deceleration with the utmost motor torque. Thus, we propose the use of spindle amplifier with the tracking delay compensation control as shown in Figure 7.

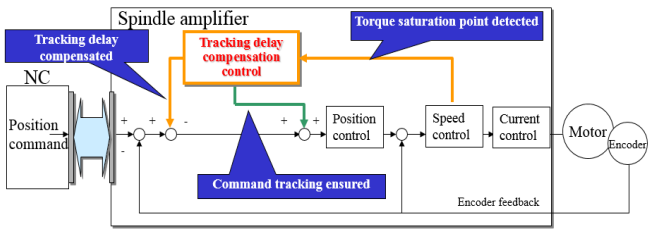


Figure 7: Tracking delay compensation control for consecutive position control.

3.2 Tracking Delay Compensation Control

Figure 8 is the block diagram to show the details of the control in the spindle amplifier which performs the tracking delay compensation control we propose.

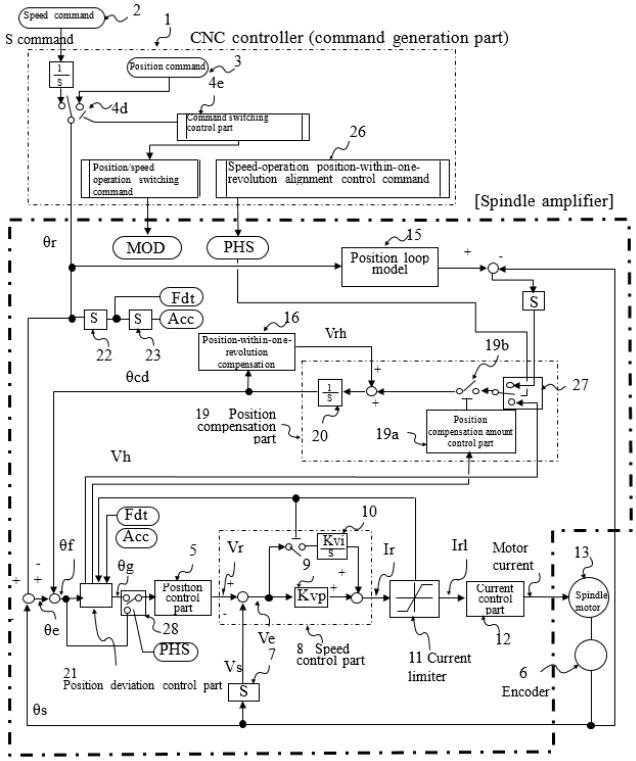


Figure 8: Tracking delay compensation.

A CNC controller 1 gives speed command signals as output to the spindle during ordinary turning, machining, or milling operations. For tracking delay compensation control, speed command signals are integrated to generate position commands. On the other hand, position command signals are output according to the spindle-end rotation angle during synchronous tapping or C-axis control. In the CNC controller 1, the position of a switch 4d is changed according to the change in the machining mode, and a position command signal θr for the spindle amplifier is generated. Meanwhile, a command switching control part 4e outputs a position/speed operation switching command MOD, which contains data to show whether position control operation or speed control operation is performed at the time of switching.

A position deviation signal θe represents the difference between the position command signal θr output by the CNC controller and a position signal θs generated by a motor-end or spindle-end encoder. After a compensation position deviation amount θcd generated by an integrator 20 is subtracted, the position deviation signal θe is input as a deviation input signal θf to a deviation control part 21. Meanwhile, the position command signal θr is converted to a command speed signal Fdt by a differentiator 22, and then

converted to a command acceleration signal Acc by a differentiator 23. The position/speed operation switching command MOD, command speed signal Fdt, command acceleration signal Acc, and deviation input signal θf are all input to the position deviation control part 21. After the predetermined operation is completed, the position deviation control part 21 outputs a deviation limiting output value θg to a position control part 5. The predetermined operation of the position deviation control part 21 will be detailed later.

The value θg output from the position deviation control part 21 is input to the position control part 5, and the position control part 5 converts the value to a speed command and outputs a speed command calculation signal Vr. Then, the speed command calculation signal Vr and a speed deviation signal Ve sent by a differentiator 7, which represents the difference from a differential of the position signal θs detected by an encoder 6, are input to a speed control part 8.

The current limiter 11 limits the current command value to a maximum current value allowed to be output by a current control part 12. The current control part 12 controls a current of a motor 13 based on a current limit value output by the current limiter 11.

While a current is limited by the current limiter 11, the current limiter 11 outputs a current limit command Il to the integral speed controller 10 to stop the integral action. The integral speed controller 10 stops the integral action to avoid unnecessary integral action of the speed deviation signal Ve while the current is limited, and to prevent overshooting from occurring for the speed command value after the current limit is released. The current limiter 11 outputs the current limit command Il also to the position deviation control part 21.

Now, the position deviation control part 21 is described in details. As shown below, the position deviation control part 21 gives a command based on the position/speed operation switching command MOD, command speed signal Fdt, and command acceleration signal Acc to a position compensation amount control part 19a of a position compensation part 19 to turn on a switch 19b according to the predetermined condition. The control part 21 also outputs an input/output deviation signal Vh, and outputs the position deviation control output value θg to the position control part 5.

Figure 9 shows the flowchart of the operation carried out by the position deviation control part 21. In the position deviation control part 21, while the current limit command Il is given and the operation mode is speed control during which the position/speed operation switching command (MOD) does not require absolute position tracking, the position deviation amount control part 19a in Figure 8 turns on the switch 19b. Then, position deviation amount θcd is created. The compensation position deviation amount θcd is subtracted from the position deviation signal θe.

Therefore, when the current command value reaches the limit in the motor controller due to saturation of the motor output voltage during acceleration/deceleration of the motor, insufficient torque for the commanded acceleration, or other reason, it is possible to prevent the gap between the speed command calculation signal Vr (output from the position control part 5) and the actual motor speed Vs from increasing even if the position command signal θr value (converted from the speed command signal) is too large while the speed command 2 is selected with the switch 4d. Consequently, it is possible to reduce the delay in returning to position control when the current limit is removed.

A position-within-one-revolution compensation control part 16 is present on the output side of the position compensation part 19. The position-within-one-revolution compensation control part 16 normalizes the compensation position deviation amount θcd output by the position compensation part 19, and calculates a position-within-one-motor-revolution deviation signal Vrh (the data for two or more revolutions are discarded and the motor deviation amount within one revolution (difference between the position command and the actual motor position) is calculated). When it is confirmed that the current command

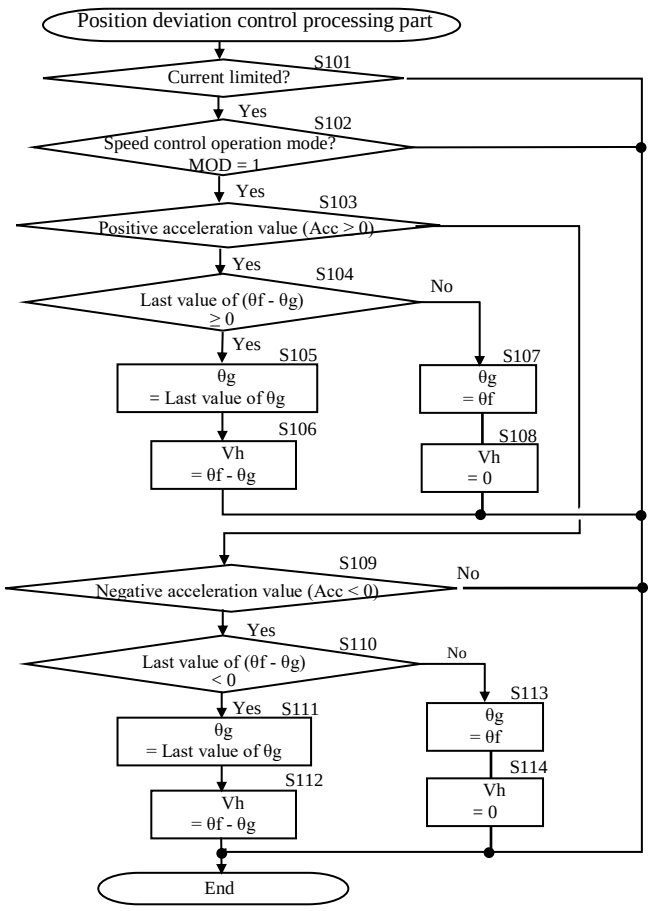


Figure 9: Operation flowchart in the position.

is within the current limit after the current limit is removed, a position-within-one-revolution compensation amount Vrh is calculated so that the position within one motor revolution deviation becomes zero. Vrh is added to the integrator 20 in the position compensation part 19.

Figure 10 (b) shows the acceleration/deceleration slope profiles of the position-within-one-revolution deviation signal Vrh. The horizontal axis represents the speed, the vertical axis the slope (acceleration), the solid line the case 1 profile, the chain line the case 2 profile, and the dash-dotted line the motor torque characteristics. As shown in Figure 10 (a), the maximum speed in the Vrh compensation pattern is determined at a certain ratio γ (e.g. 10%) of the speed feedback value at the start point of actual compensation.

The slope profile (acceleration) to the maximum speed in the Vrh compensation pattern is determined according to the motor output torque characteristics as shown by the dash-dotted line in Figure 9 (b). The slope profile line may be straight as shown by the solid line (case 1) in Figure 10 (b) by leaving a margin for the motor output torque characteristics if the controller's processing time and the memory capacity allow. Otherwise, the profile may be set in stages as shown by the chain line (case 2). Thus, the compensation is performed in a stable and fast way for the position within one revolution after the current limit is removed.

In the CNC controller 1 in Figure 8, when a command PHS, which shows that compensation is not required for the position within one motor revolution, is input, the switch 27 is switched to connect to the opposite side of the Vh signal. And the switch 28 is also switched and the deviation input signal θf is directly input to the position control part 5 by a route passing through the position deviation control part 21.

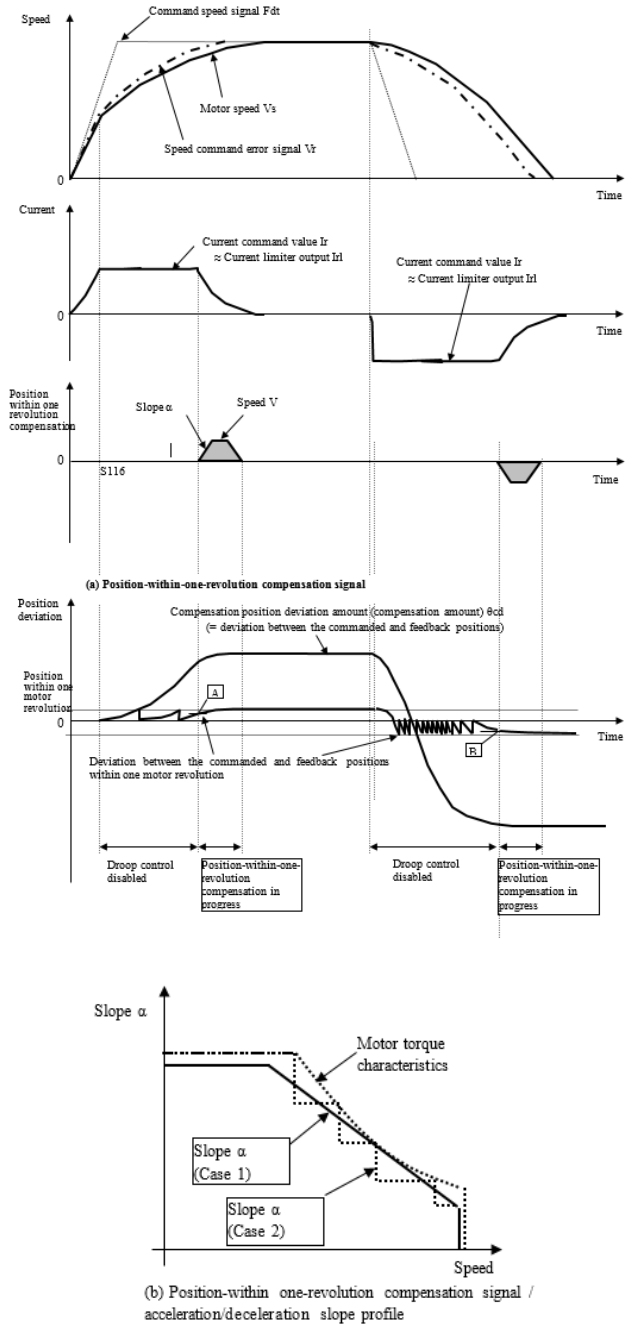


Figure 10: Tracking delay compensation control signal waveform.

Thus, when compensation is not required for the position within one motor revolution, acceleration/deceleration time can be minimized according to the motor output torque.

Figure 10 also shows waveforms of the operation signals when the proposed approach is taken. In the topmost graph, the horizontal axis represents the time, the vertical axis the speed, the chain line the speed command signal Fdt, the dash-dotted line the speed command calculation signal Vr, and the solid line the motor speed Vs. In the second graph from the top, the horizontal axis represents the time, the vertical axis the current, and the solid line the current command value. The third graph shows the position-within-one-revolution compensation signal Vrh. The horizontal axis represents the time, and the vertical axis the compensation amount. In the fourth graph, the horizontal axis represents the time, the vertical axis the position deviation, and the solid lines represent the compensation position deviation amount θcd and the deviation of the position within one motor revolution. In our approach, even if the motor cannot be fully accelerated due to the current limit and the motor speed Vs deviates significantly from a command speed signal Vrv output by the command generation part 1, the position deviation control part 21 limits the position deviation control part output value θg under certain conditions to prevent the speed deviation signal Ve, which represents the gap between the speed command calculation signal Vr output by the position control part 5 and the actual motor speed Vs, from increasing beyond the predetermined level. Thus, transition to position compensation is performed faster when the motor output torque characteristics are restored and the current limit is removed, preventing overshooting from occurring for speed or position control.

Furthermore, after the current limit is removed, the position-within-one-revolution compensation amount Vrh is added to the integrator 20 in the position compensation part 19 to enable consecutive control of the spindle position within one revolution to follow the command. The compensation position deviation amount θcd is reduced during acceleration and increased during deceleration to make the position within one revolution zero. Therefore, it takes less time to determine the compensation amount and perform the compensation for the position within one revolution.

4 VERIFICATION OF THE PROPOSED APPROACH BY SIMULATION

The following shows the result of verification by simulation of the effect of the proposed tracking delay compensation control.

Figure 11 shows the acceleration/deceleration waveform for an ordinary position loop without the tracking delay compensation control we propose. When a program is built with a time constant that would allow the torque to exceed the maximum motor torque. The current command behavior becomes unstable and causes reciprocation between upper

and lower current limits. This phenomenon causes increase in mechanical impact and significant damage on the machine.

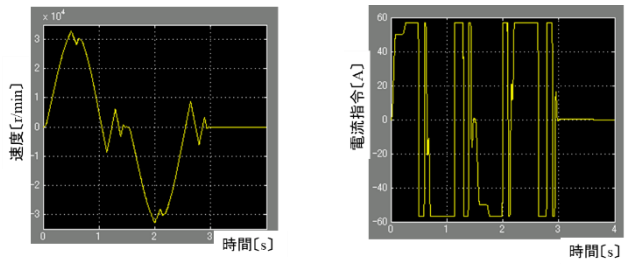


Figure 11: Acceleration/deceleration waveform without tracking delay compensation control.

Figure 12 shows the acceleration/deceleration waveform with the tracking delay compensation control we propose. Since the motor torque is restricted against the command, the current is limited at the maximum current in the amplifier. Nevertheless overshooting does not occur and the control loop remains stable.

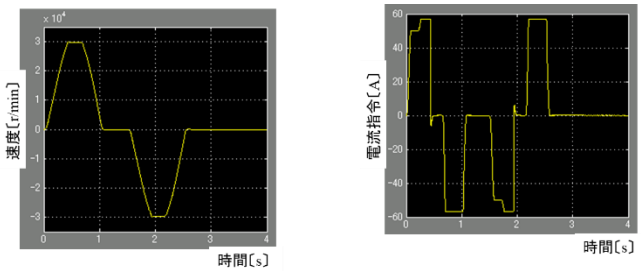


Figure 12: Acceleration/deceleration waveform without tracking delay compensation control.

5 EVALUATION

5.1 Reduction in the Time Required for Control Mode Switching by Performing Consecutive Position Control

The following shows the result of verification of the effect of implementing the proposed approach with an actual spindle amplifier.

The system for verification is a multi-tasking machine with two main spindles for turning and a tool spindle for milling. Each main spindle can be used for machining of different workpieces, or a workpiece can be pre-cut by spindle 1, re-chucked by spindle 2 for machining without set-up change operation by an operator. Spindle 1 and spindle 2 can be also used to chuck one workpiece for heavy cutting.

Table 1 shows the time reduction result for various mode of spindle control.

Table 1: Effect of time reduction for operation of the multi-task machining program

New function	Description	Effect		
			Previous amplifier	New amplifier
Spindle consecutive position control	Seamless transition of the operation pattern for switching to C-axis or spindle synchronous control	**Switching to C-axis control**	**0.8 to 1 s**	**0 s**
		Switching to spindle synchronous control	**11.2 s**	**0 s**

6 CONCLUSION

In the past, a focus is put on high-speed revolutions of the tool for spindle motors and spindle amplifiers in machine tools. Therefore, the critical issues are to generate the maximum torque to reduce the acceleration time to reach high speed revolutions, and to generate high power for heavy cutting.

However, in the recent progress in pursuing high productivity for cutting machining, multi-tasking machines for both turning and milling have become more and more dominant. As a consequence, position control such as C-axis control has become essential for spindles in increasing number of cases and how to reduce the switching time has become a controversial issue.

However, the other ways except for switching had not existed before our proposal. Because no other way but our proposal couldn't solve the overshooting issue of spindle's position loop during acceleration/deceleration with the utmost motor torque.

In this research, we proposed consecutive position control in spindle amplifiers in the same way as in servo amplifiers to address the issue, and verified its validity. This control method has been adopted in our products, which differentiates our products from previous spindle amplifiers and controllers. Furthermore, we are in the process of promoting standardization of amplifiers for motors driving machine tools to remove the barriers between different types of amplifiers.

REFERENCES

[1]Perspective of numerical control for machine tool, Journal of the Japan Society for Precision Engineering, 1985, Vol.24.No.286

[2]Recent servo controller for NC machine tool Journal of the Japan Society for Precision Engineering, 1993, Vol.59.No.9

[3]All digital high performance controller for spindle motor in CNC machine tool IEEE International Electric Machines and Drives Conference Record, 1997, MC2-2.1

[4]Current State and Outlook of the Multi-tasking Machine, Journal of the Japan Society for Precision Engineering, 2012, Vol.78.No.9

[5] Driving system MDS-D/DH series for M700, Mitsubishi Electric Advance, Dec.2006, Vol.116

Failure prediction considering interaction between servo motors of FA equipments

Masanori Haga†, Kazuhiko Tsutsui‡, Katsuhiko Kaji*, Katsuhiro Naito*, Tadanori Mizuno*, Naoya Chujo*

†Graduate School of Informatics, Aichi Institute of Technology, Japan
‡Mitsubishi Electric Co., Ltd., Japan
* Aichi Institute of Technology, Japan
{Tsutsui. Kazuhiko}@ay.MitsubishiElectric.co.jp
{b19723bb, kaji, tmizuno, ny-chujo}@aitech.ac.jp
{naito}@pluslab.org

Abstract - The failure of FA equipment used in the factory is important because it affects the productivity. Recently, predictive maintenance has been considered as a maintenance system for FA equipment. The predictive maintenance is a system in which sensors are attached to an equipment. The failure is predicted using measured values, and waste of conventional preventive maintenance can be omitted, because the condition of the equipment is individually managed. However, there are problems such as that the number of sensors to be installed is large and that it is necessary to secure the space to install the sensors.

It is considered that the interaction such as vibration is transmitted between the links of the parallel link robot. It was considered that the information which led to the predictive maintenance of other link and the entire equipment could be known from the sensor of one link by using interaction.

In order to confirm this, an experimental apparatus equipped with a parallel link was produced experimentally, and measurement experiment was carried out by applying a load corresponding to a failure.

Principal component analysis was carried out on the measured data, and the result which supported the detection was obtained.

Keywords: Predictive Maintenance, Factory Automation, Parallel Link, Principal Component Analysis

1 INTRODUCTION

Today's factory automation(FA) systems support a wide range of industrial products, from electronic devices such as mobile phones to transportation equipment such as automobiles. Proper maintenance of FA systems is required to ensure the quality of these products. In general, it is said that the maintenance cost of an FA system is 15 to 60% of the manufacturing cost of the product, and thus is not insignificant[1].

There are two types of maintenance methods for FA equipment. The first is preventive maintenance which involves regularly performing maintenance, regardless of the state of the equipment. Therefore, unnecessary maintenance is performed on the FA equipment, which increases maintenance costs. The second type of maintenance is predictive maintenance, in which maintenance is performed according to the predicted state of the FA equipment. The state of the FA equipment is measured by sensors, and future failure times are predicted. It is said that the maintenance cost can be reduced by conducting maintenance based on failure prediction. The International Air Transport Association(IATA) has estimated that the cost of maintenance would be reduced by 15 to 20% if aircraft were to be subjected to predictive maintenance[2]. Failure prediction is required to perform predictive maintenance. In addition to the installation cost of a sensor and physical space in a system, including wiring, must also be provided for the sensor[3]. In addition, a communication channel must be secured in order to upload real-time measurement data to the cloud.

The purpose of the present paper is to predict the failure of FA equipment. FA equipment consists of a combination of several servo motors, links, and other mechanical components. By attaching sensors to the components in a system, it is possible to accurately measure mutations that lead to component failure. We herein examine the variability of data due to interactions.

In the present paper, we detect failure for predictive maintenance using the interaction between mechanical parts. Defective components in FA equipment are detected based on changes in the sensor data. We examine the variability of data due to interactions, while increasing the load on the equipment until failure. We use principal component analysis (PCA) for feature extraction in the failure prediction.

The remainder of the present paper is organized as follows. Section 2 introduces related research. Section 3 presents the method for failure detection. Section 4 describes the experimental system, the experimental method, and the obtained results. Section 5 discusses the evaluation results, and a summary is presented in Section 6.

2 RELATED RESEARCH

System diagnostics can be separated into model-based diagnostics and signal-based diagnostics[4]. Model-based diagnosis is used when the theoretical modeling of a target system is straightforward. A deterministic model is created using equations to represent the actual system, and the diagnosis is performed by comparing the output of the system with the output of the model.

Signal-based diagnosis is used when theoretical modeling of a target system is difficult. In this case, we use a model derived from measurements. We extract the characteristics for the normal and abnormal operating conditions of the equipment from the measurements and use these characteristics to

diagnose faults.

In recent years, methods for signal-based diagnosis using machine learning techniques have been introduced. With the increasing accuracy of sensors and developments in the field of AI, the accuracy of failure prediction is improving. There are also techniques to perform signal-based diagnosis under dynamic conditions, compared to static conditions alone [5].

FA equipment is generally reliable and resistant to failure. However, in real environments, it is difficult to apply supervised learning because of the small accumulation of failure data. Therefore, it is common to use unsupervised machine learning for FA equipment.

According to a review paper on FA equipment diagnosis by unsupervised machine learning, methods using PCA provide the best results[6].

However, there is a problem in signal-based diagnosis in that signal based diagnostics require high computational power for recording and process large amounts of measurement data. Therefore, it is necessary to apply these methods using cloud computing[7]. In the case of FA equipment, it is necessary to secure a large capacity communication line.

For this reason, it is desirable that the computer on the edge side be able to diagnose faults in the FA equipment. In addition, it is desirable that the number of sensors be reduced, when edge computing is used. There are interactions between components such as the resonance between the components in the FA equipment, and the status of the entire FA system can be diagnosed using a small number of sensors with appropriate signal processing. However, as far as we know, methods for predicting failure using the interaction between parallel links in FA equipment have not been studied.

3 FAILURE DETECTION OF FA EQUIPMENT WITH PARALLEL LINKS

In the present paper, we verify a method for detecting failure by signal-based diagnosis for FA equipment with parallel links.

Parallel-link robots[8] consist of multiple link mechanisms of the same type that are arranged in parallel and operate synchronously. Robots using parallel links use many more shared control components compared to a single-arm robot. Therefore, the manufacturing cost is low, and the mechanism is simple and easy to maintain. For this reason, parallel links have been widely used.

Fault detection is performed by the parallel link mechanism with the installed sensor. Since each link is combined to form a single mechanism, the movement of a link is affected by other links. Therefore, if there is an abnormality in the operation of a certain link, the behavior of other links is expected to be affected.

In the present paper, we consider the interaction between parallel links. By considering this interaction, fault detection is performed by detecting abnormalities in other link mechanisms using only the measurement data of the sensors from a single link. In the next section we discuss how to verify the degree of abnormal behavior from other links that can be detected.

4 EXPERIMENTAL METHOD

In this section, we describe the experimental apparatus with a parallel link as well as the procedure for using the experimental apparatus. An outline of the experiment is as follows.

- Applying the load which assumes an increase in the friction load at a specific link, and measuring the data.
- Analysis of changes due to increased load
- Comparison of data between loaded and non-loaded links to verify the possibility of failure detection.

4.1 Experimental Apparatus

The experimental apparatus is shown in Fig. 1. The apparatus is a system to detect failure time based on data obtained from a servo motor.

The system consists of a PC and parallel link mechanisms using two servo motors. The PC transfers control to the servo motors and the servo motors transmit the data to the PC. Six kinds of data can be obtained from the servo motors.

The sensor data consist of temperature, current, voltage, rotational angle, rotational speed, and rotation time that can be obtained from the servo motor. The data acquisition frequency is 10 Hz. The temperature data measurements of the servo motor are affected by the temperature of the room. Therefore this temperature is stabilized in the range of ± 1 °C using air conditioning. This method prevents temperature change due to factors other than the heat generated by the operation of the servo motor.

The equipment is approximatelt 15 cm and is smaller than actual FA equipment. The specifications of the servo motor are listed in Table 1.

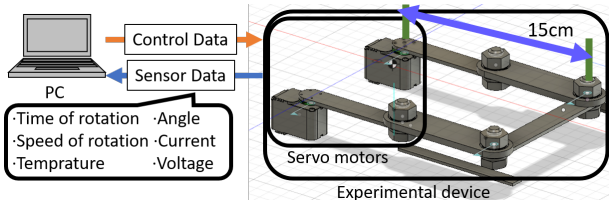


Figure 1: Experiment system

Table 1: Specifications of the servo motor (RS302CD)

Torque (during operation 7.4 V)	5.0 kgf·cm
Current consumption (when stopped)	40 mA
Current consumption (when moving)	125 mA
Working voltage	7.2–7.4 V
Movable angle	150°
Temperature limit	0–40°C
Communication speed	max 460.8 kbps

In the experiment, by operating the experimental apparatus which is similar to FA equipment, we obtain the sensor data necessary for failure detection by increasing the load of the servo motor. In particular, we increase the load on the link

component in order to create abnormal operating conditions. The experimental apparatus is shown in Fig. 2.

The joints connect the links of the link components. The guide rails limit the movement of the drive unit. The drive unit is perpendicular to the guide rail.

The drive unit moves backward and forward by rotating servo motors A (M_A) and B (M_B) outward. The weights are placed either at Joint A (J_A) or Joint B (J_B).

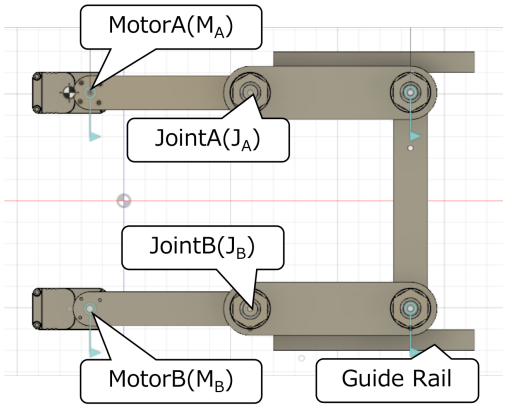


Figure 2: Parallel links

4.2 Experimental Procedure

Experiments are conducted to increase the friction between the links. First, we record the sensor output for 20 minutes without placing a load on the system. Second, weights are fixed to J_A for a 30-second interval. Next, we record the sensor output for another 20 minutes under a 70 g load. Weights are fixed to J_A, again. Finally, we record the sensor output for 20 minutes under a 130 g load.

This experiment is carried out in order to verify whether the load increase can be observed by only the measurements from M_B when placing a weight on J_A.

5 RESULTS

We first describe the observed interactions. Next, we describe the results of PCA using 12-dimensional data. Finally, we describe the results of PCA using six-dimensional data.

The data obtained from each servo motor are shown in Figs. 3, 4, 5, 6, 7, 8, 9, and 10.

5.1 Changes in Rotational Angle

Figs. 3 and 4 show the rotational angles of the servo motors. In the rotational angle data, M_A indicates that the movement of the horn is faster than usual due to the loading. Fig. 5 shows an enlarged view of the angle data for M_B. The region of the rotational angle data containing the abnormal readings is enclosed in a black circle. Fig. 5 shows an enlarged view of the black circle. In M_B, there was an interaction in which the timing of the action was shifted by M_A.

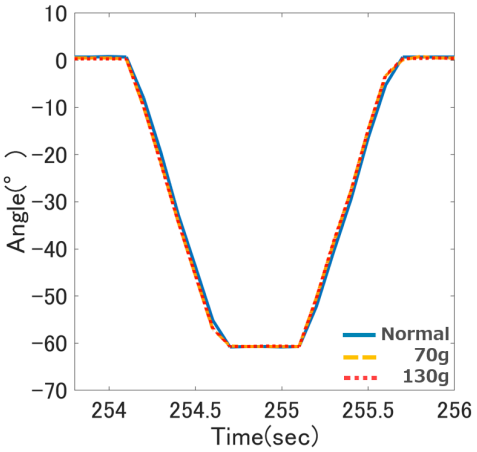


Figure 3: Angle data for Motor A

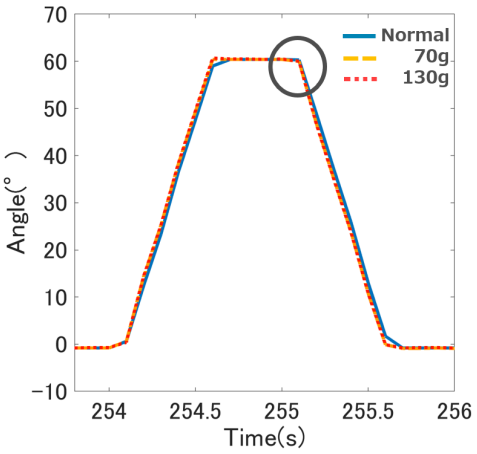


Figure 4: Angle data for Motor B

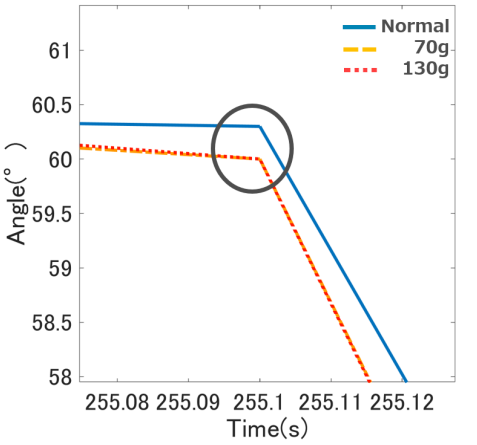


Figure 5: Enlarged view of the angle data for Motor B

5.2 Changes in Rotation Time

Figs. 6 and 7 show the rotation times of the servo motors. Rotation time indicates the elapsed time from the start of the movement of the servo horn, and the value is retained until the next movement after arrival at the target angle. The abnormal measurements in the rotation time data are enclosed in a black

circle. Fig. 8 shows an enlarged view of the area indicated by the black circle. Here, M_A indicates that the time of movement is clearly earlier than that for the case without a load. Fig. 8 shows an enlarged view of the M_B rotation time data. In addition, M_B is also moving slightly faster than before the load was added.

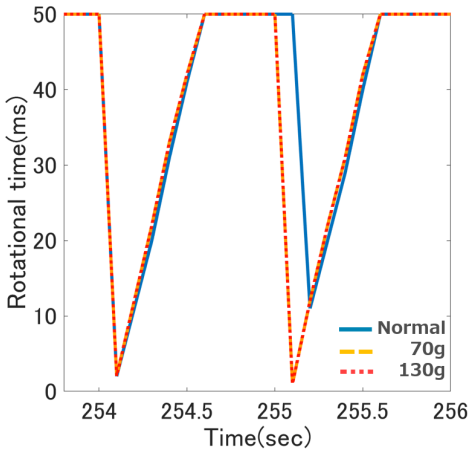


Figure 6: Rotation time data for Motor A

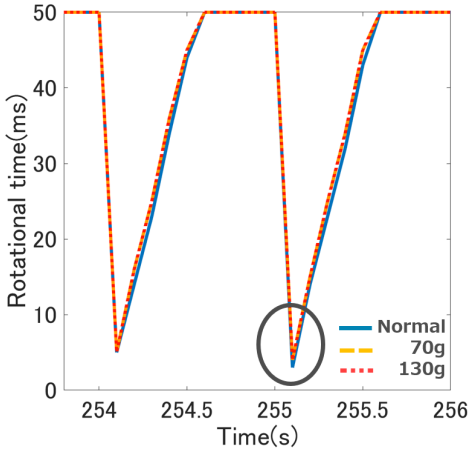


Figure 7: Rotation time data for Motor B

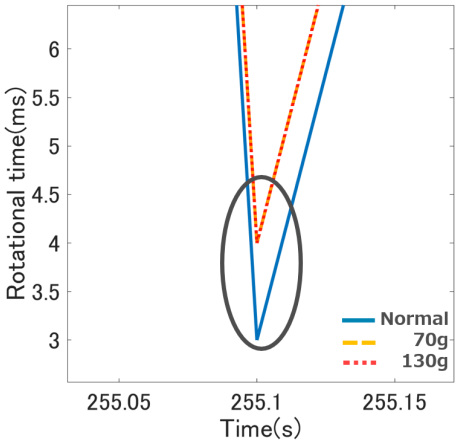


Figure 8: Enlarged view of the rotation time data for Motor B

5.3 Changes in Voltage

Figs. 9 and 10 show the voltages of the servo motors. The three sets of measurements for M_A show that the tendencies in the voltage change do not agree. The M_A voltage under the 130-g load is more stable than that under the 70-g load, probably because the play in the experimental apparatus was suppressed by the weight. The abnormal voltage data are enclosed by a black ellipse. The voltage data for M_B show a small variation, but an abnormal voltage drop was observed due to the interaction.

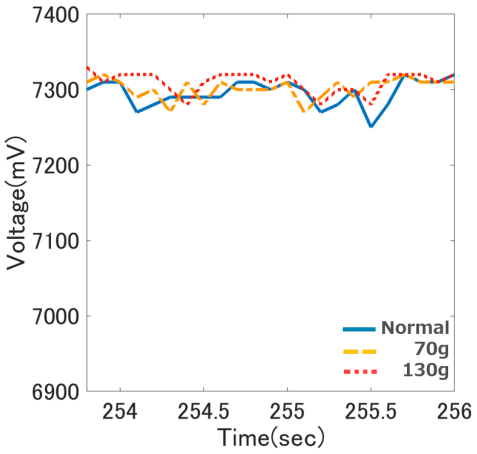


Figure 9: Voltage data for Motor A

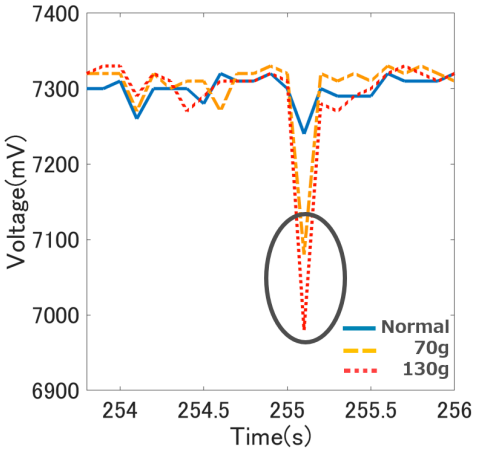


Figure 10: Voltage data for Motor B

5.4 Results of PCA

PCA was performed using 12-dimensional data from both motors.

Figs. 11 and 12 show the 12-dimensional data distribution without and with loads, respectively. The number of classes could be observed to change as the load increased. Classes are formed by grouping points for each operation of the experimental equipment. A given point moves between multiple

classes after one movement of the experimental equipment, before returning to the initial class. Classes P_1 through P_4 are highlighted in Fig. 11. Other classes are moving between assigned numbers. The classes are changing as the load increases. In particular, the change of the P_3 class is remarkable.

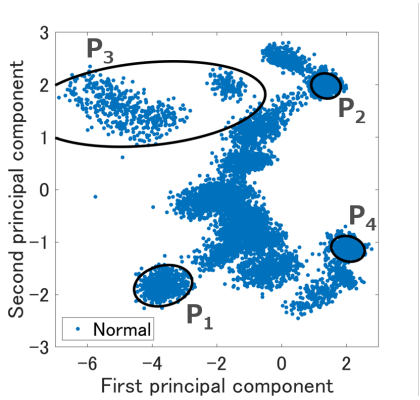


Figure 11: 12-dimensional data distribution: without load

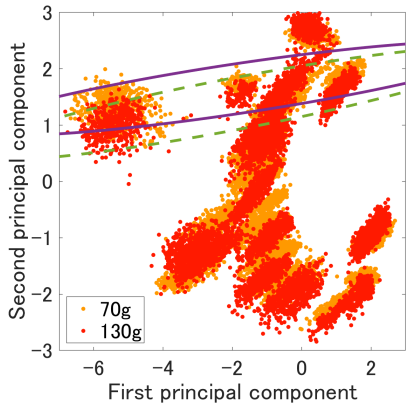


Figure 12: 12-dimensional data distribution: with load

5.5 Failure Detection Using Interactions

PCA was performed using six dimensional data for each servo motor. Figs. 13 and 14 show the results of PCA on M_A and indicate the data distribution without and with loads, respectively. Since M_A was directly loaded, the transition in the entire class from right to left was remarkable.

Figs. 15 and 16 show the results of PCA for M_B and indicate the data distributions without and with load, respectively. A transition was observed in the class. From this result, it is considered that abnormality in the equipment operation can be detected based on only PCA of the data of M_B without applying the load directly.

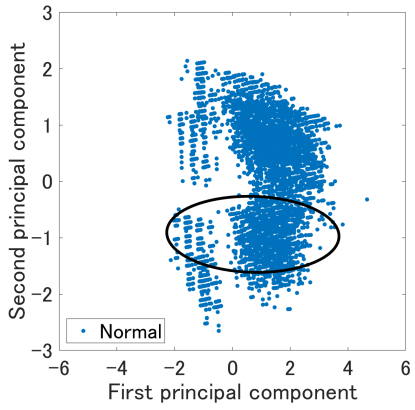


Figure 13: Six-dimensional data distribution on Motor A: without load

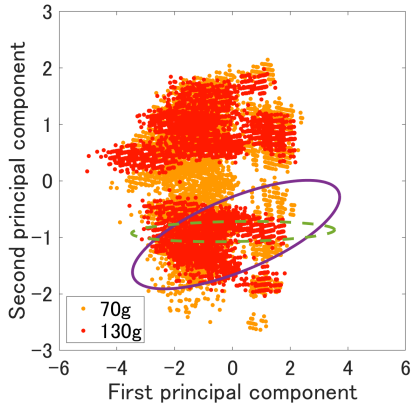


Figure 14: Six-dimensional data distribution on Motor A: with load

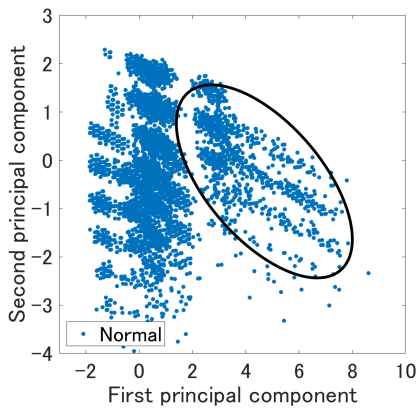


Figure 15: Six-dimensional data distribution on Motor B: without load

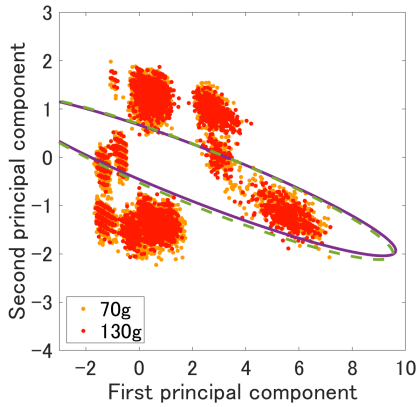


Figure 16: Six-dimensional data distribution on Motor B: with load

6 DISCUSSION

In the present paper, we examined possibility of fault detection for FA equipment with parallel links using the interaction between links.

An experimental apparatus for monitoring parallel links using two servo motors was developed. An experiment in which a load was placed on one link was carried out in order to determine whether an increase in friction could be detected from the other link. Based on the results of PCA of 12-dimensional data from the servo motors, it was confirmed that multiple data classes changed when the load was increased.

Based on the results of the PCA of six-dimensional data obtained from one servo motor, it was possible to observe changes in the data classes when increasing the load not only in the servo motor with the load but also in the servo motor without the load. This indicates that it is possible to perform fault detection based on the interaction between parallel links of FA equipment.

We thought that it is possible to predict failure by approximating the distribution of data classes with a model and estimating the speed of their movement and rotation.

7 CONCLUSION

In the present paper, we examined possibility of fault detection for FA equipment assuming there exists an interaction between parallel links.

An experimental apparatus with a parallel linkage was developed, and a load was applied to one side of the link in order to test the assumption that this would increase the friction of the link. The sensor output data were then measured.

As a result of PCA of the time series data, confirmed that multiple data classes changed as the load increased. In the analysis of the measurements for each servo motor, it was possible to observe the change in the data classes under increasing loads, not only in the servo motor with the load but also in the servo motor without the load. This indicates the possibility of fault detection based on observation of the interaction between links.

Future issues include the verifying possibility of failure prediction by analyzing approximated data class.

REFERENCES

[1] R. Keith Mobley, An Introduction to Predictive Maintenance, c. 1, Butterworth-Heinemann, 2002.

[2] IATA'S Maintenance Cost Task Force, AirLine Maintenance Cost Executive Commentary, 2016.

[3] Akihiro Yamashita, Hiroshi Mineno, Tadanori Mizuno, Distributed Remote Input/Output Control Method in Real Time Processing for CNC, International Journal of Informatics Society, Vol. 8, No. 2, 2016.

[4] Zhiwei Gao, Carlo Cecati, Steven X. Ding, A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part I: Fault Diagnosis With Model-Based and Signal-Based Approaches, IEEE Transactions on Industrial Electronics, Vol. 62, No. 6, 2015.

[5] Bo Luo, Haoting Wang, Bin Li, Fangyu Peng, Early Fault Detection of Machine Tools Based on Deep Learning and Dynamic Identification, IEEE Transactions on Industrial Electronics, Vol. 66, No. 1, 2019.

[6] Nagdev Amruthnath, Tarun Gupta, A Research Study on Unsupervised Machine Learning Algorithms for Early Fault Detection in Predictive Maintenance, 5th International Conference on Industrial Engineering and Applications, 2018.

[7] Sachchidanand Singh, Optimize cloud computations using edge computing 2017 International Conference on Big Data, IoT and Data Science (BID), 2017.

[8] Jean-Pierre Merlet, Parallel Robots (Solid Mechanics and Its Applications),pp. 12-17, Springer Netherlands, 2000.