IWIN2022

International Workshop on Informatics

Proceedings of
International Workshop on Informatics

August 31-September 3, 2022
Nachi-Katsuura

Sponsored by Informatics Society

IWIN2022

International Workshop on Informatics

Proceedings of
International Workshop on Informatics

August 31-September 3, 2022
Nachi-Katsuura

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:
978-4-902523-49-2
Printed in Japan

Table of Contents

Message from the General Chairs


It is our great pleasure to welcome all of you to Nachi-Katsuura, Wakayama, Japan for the 16th International Workshop on Informatics (IWIN 2022). This workshop has been held annually by the Informatics Society. Since 2007, the workshops were held in Naples in Italy, Wien in Austria, Hawaii in the USA, Edinburgh in Scotland, Venice in Italy, Chamonix in France, Stockholm in Sweden, Prague in Czech Republic, Amsterdam in Netherlands, Riga in Latvia, Zagreb in Croatia, Salzburg in Austria, Hamburg in Germany, Wakayama in Japan (virtually), and Fukui in Japan (virtually), respectively.

In IWIN 2022, 25 papers were accepted after peer reviewing by the program committee. Based on the papers, eight technical sessions were organized in a single-track format, which highlighted the latest research results in the areas such as Artificial Intelligence (AI), Systems and Services, Multimedia, Disaster Mitigation, Social Network and Data Analysis, Communication, Application, and Internet of Things (IoT) and Network. IWIN2022 will also welcome one keynote speaker: Mr. Takeshi Saito, general manager of Information and Communication Platform Laboratories of Corporate R&D Center, Toshiba Corp. We really appreciate his 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 to make the workshop a great success. We are looking forward to seeing you all in the workshop. We hope you enjoy IWIN 2022.

August 2022

Tomoyuki Yashiro
Hideki Goromaru

Organizing Committee

General Co-Chairs

Tomoyuki Yashiro (Chiba Institute of Technology, Japan)
Hideki Goromaru (Chiba Institute of Technology, Japan)

Steering Committee

Hitoshi Aida (The University of Tokyo, Japan)
Toru Hasegawa (Osaka University, Japan)
Teruo Higashino (Kyoto Tachibana University, Japan)
Tadanori Mizuno (Aichi Institute of Technology, Japan)
Jun Munemori (The Open University of Japan, Japan)
Yuko Murayama (Tsuda University, Japan)
Ken-ichi Okada (Keio University, Japan)
Norio Shiratori (Chuo University / Tohoku University, Japan)
Osamu Takahashi (Future University Hakodate, Japan)

Program Chair

Kei Hiroi (Kyoto University, Japan)

Financial Chair

Tomoya Kitani (Shizuoka University, Japan)

Publicity Chair

Yoshitaka Nakamura (Kyoto Tachibana University, Japan)

Program Committee

Akihiro Hayashi
(Shizuoka Institute of Science and Technology, Japan)
Akihito Hiromori (Osaka University, Japan)
Fumiaki Sato (Toho University, Japan)
Hideyuki Takahashi (Tohoku Gakuin University, Japan)
Hironobu Abe (Tokyo Denki University, Japan)
Hiroshi Inamura (Future University Hakodate, Japan)
Hiroshi Mineno (Shizuoka University, Japan)
Hiroshi Sugimura
(Kanagawa Institute of Technology, Japan)

Hiroshi Yoshiura (Kyoto Tachibana University, Japan)
Kanae Matsui (Tokyo Denki University, Japan)
Katsuhiko Kaji (Aichi Institute of Technology, Japan)
Kazuyuki Iso (NTT, Japan)
Kei Utsugi (Hitachi, Ltd, Japan)
Keiichi Abe (Kanagawa Institute of Technology, Japan)
Ken Ohta (NTT DOCOMO, Japan)
Kozo Okano (Shinshu University, Japan)
Makoto Imamura (Tokai University, Japan)
Mari Ueda (Kanagawa Institute of Technology, Japan)
Masaji Katagiri (iU, Japan)
Masakatsu Nishigaki (Shizuoka University, Japan)
Masaki Endo (Polytechnic University, Japan)
Masaki Nagata (Shizuoka University, Japan)
Michiko Oba (Future University Hakodate, Japan)
Mikiko Sode
(Kanazawa Institute of Technology, Japan)
Minoru Kobayashi (Meiji University, Japan)
Naoya Chujo (Aichi Institute of Technology, Japan)
Orie Abe
(Information Communication Media Laboratory, Japan)
Ryozo Kiyohara
(Kanagawa Institute of Technology, Japan)
Satoru Matsumoto (Osaka University, Japan)
Shigemi Ishida (Future University Hakodate, Japan)
Shigeyoshi Ohno (Polytechnic University, Japan)
Shinichiro Mori (Chiba Institute of Technology, Japan)
Shinji Kitagami
(Fukui University of Technology, Japan)
Takaya Yuizono
(Japan Advanced Institute of Science and Technology)
Takayasu Yamaguchi
(Akita Prefectural University, Japan)
Takuya Yoshihiro (Wakayama University, Japan)
Tetsuya Shigeyasu
(Prefectural University of Hiroshima, Japan)
Tetsuya Yokotani
(Kanazawa Institute of Technology, Japan)
Tomoki Yoshihisa (Osaka University, Japan)
Tomoo Inoue (University of Tsukuba, Japan)
Tsukasa Kudo
(Shizuoka Institute of Science and Technology, Japan)
Yoshia Saito (Iwate Prefectural University, Japan)
Yoshiaki Terashima (Soka University, Japan)
Yoshinobu Kawabe
(Aichi Institute of Technology, Japan)
Yu Enokibori (Nagoya University, Japan)
Yuichi Bannai
(Kanagawa Institute of Technology, Japan)
Yuichi Tokunaga
(Kanazawa Institute of Technology, Japan)
Yuki Koizumi (Osaka University, Japan)
Yukimasa Nagai
(Mitsubishi Electric Corporation, Japan)
Yusuke Gotoh (Okayama University, Japan)
Yusuke Ichikawa
(NTT Service Evolution Laboratories, Japan)

Session 1:
Artificial Intelligence
(Chair: Kei Hiroi)

Improve Measuring Suspiciousness of Bugs in Spectrum-Based Fault Localization With Deep Learning

Hitoshi Kiryu†, Shinpei Ogata‡, and Kozo Okano‡

†Graduate School of Engineering, Shinshu University, Japan
†21w2025g@shinshu-u.ac.jp
‡Faculty of Engineering, Shinshu University, Japan
‡{ogata, okano}@cs.shinshu-u.ac.jp

Abstract - Localizing Faults is integral for debugging in developing software. Spectrum-Based Fault Localization (SBFL) is a technique to localize faults. SBFL calculates the suspiciousness score for each line in a program using code coverage of tests. Some studies apply a deep neural network to SBFL. In these studies, the suspiciousness is calculated by giving a virtual coverage to a trained network. This paper proposes a method to calculate the suspiciousness score of lines in a program from the program spectrum with deep learning. The method also provides a ranking based on the score. In calculating the suspiciousness score, We focus on the difference between the coverage of a failed test and an alteration of the coverage. As a result of the evaluation of the proposed method, it is confirmed that the effectiveness of the proposed method is comparable to Ochiai and more effective than the suspiciousness calculation method used in other studies.

Keywords: SBFL, Fault Localization, Deep Learning

1 INTRODUCTION

When a problem caused by code is found in software development or maintenance, it is necessary to localize and fix bugs. Generally, such a debugging needs a lot of time and human works. Many techniques to localize faults and fix bugs have been studied to support developers in debugging. One example of a bug-fixing technique is GenProg[1], which outputs code that passes all the test suites with a genetic algorithm. In technique to localizing bugs, Various studies [2]–[5] have been conducted. These studies proposed methods to identify statements that cause bugs by using bug reports, trace information, and visualization.

Another technique for localizing the fault is Spectrum-Based Fault Localization (SBFL). SBFL localizes faults in the source code based on the code coverage and test results of each test. It calculates the suspiciousness of each statement containing a bug and provides a ranking based on the suspiciousness. The basic idea of SBFL is that a line executed in a failed test is more likely to contain a bug, while a line executed in a successful test is less likely to contain a bug. The metrics for calculating the suspiciousness of a bug are based on four values described in Table 1.

```
     def fizzbuzz(i):
S1     if i % 15 == 0:
S2       return "FizzBuzz"
S3     elif i % 3 == 0:
S4       return "Fizz"
S5     elif i % 4 == 0:
     # correct condition is “i % 5 == 0”
S6       return "Buzz"
S7     return i
```

	i	S_1	S_2	S_3	S_4	S_5	S_6	S_7	test result
test1	15	1	1	0	0	0	0	0	0
test2	3	1	0	1	1	0	0	0	0
test3	4	1	0	1	0	1	1	0	1
Ochiai		0.577	0.0	0.707	0.0	1.0	1.0	0.0	

Figure 1: Example of SBFL

Table 1: Four Values for Calculation of Suspiciousness

e_f	Number of failed tests that execute the program element.
e_p	Number of passed tests that execute the program element.
n_f	Number of failed tests that do not execute the program element.
n_p	Number of passed tests that do not execute the program element.

$$Ochiai = \frac{e_f}{\sqrt{(e_f + n_f)(e_f + e_p)}} \quad \text{(Ochiai)}$$

$$Dstar(* = N) = \frac{{e_f}^N}{e_p + n_f} \quad \text{(Dstar)}$$

For example, Ochiai[6] and Dstar[7] proposed following formulas. Here, ‘N’ is the exponent variable of e_f. Suspiciousness scores for each line are calculated by these formulas, and represents the probability that each line causes a bug. The score tends to be higher if a line is executed more frequently when the test fails.

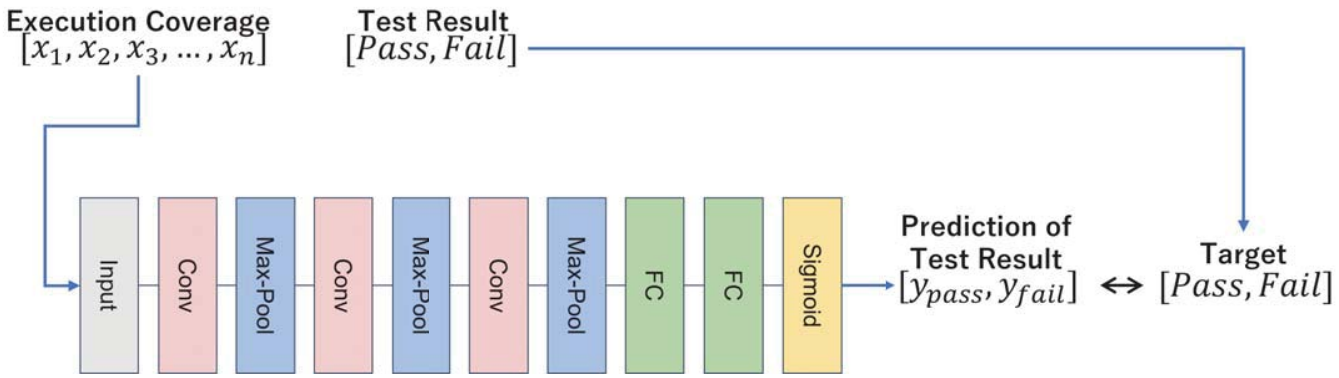


Figure 2: Architecture of the Network

Figure 1 shows an example of SBFL. The source code in the figure is a function that returns the result of FizzBuzz for a given number i as input. The fifth line of the function contains a bug. This is because the conditional statement is incorrect. The suspiciousness of each line is calculated by Ochiai based on the results and the coverage of the tests. The suspiciousness is the highest on lines fifth and sixth, which are executed only when the test fails. This indicates that the suspiciousness of the buggy lines is calculated properly. Thus, from the test results and the coverage, it is possible to identify the location of bugs.

Deep learning is a technology that has been showing results in a wide variety of fields, including image recognition and natural language processing. The field of fault localization is no exception either. Ikeda et al.[8] proposed the method that localizes a fault method with a neural network. They trained the network to predict test results from execution traces and used ablated traces to localize a fault method. In the field of SBFL, the study [9] has been conducted to compute the suspiciousness using deep learning. In this study, three architectures, RNN (Recurrent Neural Network), CNN (Convolutional Neural Network), and MLP (MultiLayer Perceptron), are proposed to compute the suspiciousness of bugs, and CNN shows the best results. In another study[10], which calculates the suspiciousness using RBF networks (Radial Basis Function Network), concluded that RBF networks are more effective than existing methods such as Ochiai in fault localization. In these studies, in order to calculate the suspiciousness, a virtual coverage which only a certain line is executed is input to the trained network. The output of the network is treated as the suspiciousness of the line have bugs, and a ranking based on the suspiciousness is provided.

This paper proposes a method to compute the suspiciousness of lines in the source code from the program spectrum with deep learning. The proposed method shall identify a single defect in the source code. Using a network trained to predict test results from the source code coverage of each test case, suspiciousness scores are measured for each line.We measure the impact of each line on bug prediction in the trained network, and provide a ranking of lines that are likely to cause bugs based on the measured impact. In measuring the impact of the network on a bug prediction, we focus on the change in bug prediction between an actual failed coverage and virtual coverages that altered the actual coverage. We consider the difference between them as the suspiciousness score.

The proposed method is better than the conventional method in measuring the suspiciousness from the network. It is confirmed that the proposed method gives a larger value of suspiciousness to the buggy line and a smaller value to the non-buggy line than the conventional measuring method. While comparing a metric Ochiai, no significant difference was confirmed between the proposed method and Ochiai.

In the following sections, Section 2 describe the proposed method. Section 3 presents the results in the evaluation experiments and Section 4 discuss the results in the experiments. Finally, we conclude in Section 5.

2 PROPOSED METHOD

This section describes the methodology of SBFL using a Convolutional Neural Network in detail.

The proposed method provides suspiciousness values that each line contains a bug in a source code and a ranking of the suspiciousness. CNN shows the best result in Zhang *et al*. [9]. That is why we use Convolutional Neural Network as the architecture of the neural network in the proposed method.

In the proposed method, first, a neural network learns about a relation between the coverage of the source code in tests and the test result. The network takes the coverage as input, and output respectively probabilities of a test result, pass and fail. In order to measure the impact on decision making about a test result in the network, a coverage of in failed test and a virtual coverage that altered the failed test coverage are input. We determine the impact based on a difference between outputs of the original coverage that fails the test and the virtual coverage. The impact on decision making is considered as the suspiciousness of bugs. The trained network cannot be applied to other source code. This

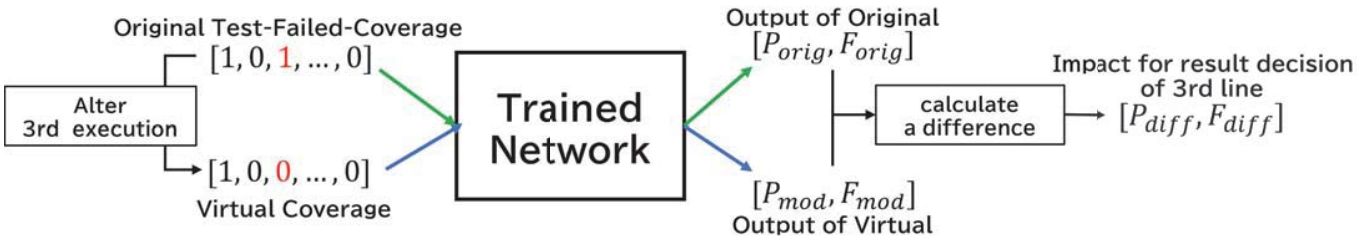


Figure 3: Example of Calculation of Suspiciousness of Third Line.

is because the network learns the relation between a execution of lines in a certain source code and test results.

2.1 Network Architecture and Training

Figure 2 shows the structure of the CNN used in the proposed method. It has two Convolutional layers, a MaxPooling layer on the output of the Convolutional layers, two Fully-Connected layers, and one output layer. The sigmoid function is used in the output layer, and ReLU is used for the other activation functions. The kernel size of the convolutional layer, the number of the channels, and the nodes of the Fully-Connected layer depend on the target program size. The network takes a test coverage as input and learns a nonlinear relationship between the coverage and the test results. The output of the network is P, and F, which represent the probability of the test passing and failing, respectively.

2.2 Extract Suspiciousness of a Bug

A virtual coverage is inputted to the network after training to measure the impact on deciding test results in the trained network based on the difference between the virtual coverage and the original coverage. Figure 3 shows an example of measuring the impact of the third line. In order to calculate the impact of the execution of the third line on the determination of test results in the network by examining the change in the output of the original failed coverage and coverage that the third line is changed to zero in original the coverage. Some studies[9], [10] treat the output of the virtual coverage in which only the line is executed as the score of the suspiciousness of a bug. We believe that the difference between the original coverage and the output of the coverage with the altered execution more strongly indicates the impact on the test results than such methods. If a line, which is changed to zero in the coverage contributes to a failure, the elimination of the execution might increase the probability that the result is passed. Since the large change in the likelihood of the network's prediction signifies the bigger impact of the execution, we consider the change to be the suspiciousness score.

The values of suspiciousness are ranked in descending order and output as a result of fault localization. The ranking presents lines and the corresponding suspiciousness scores. Lines that are not executed in failed tests are excluded from the ranking, because they are not likely to contain bugs.

3 EVALUATION EXPERIMENT

In order to evaluate the proposed method, evaluation experiments are conducted on two OSS to see if the proposed method is effective in localizing faults. The coverage used in the experiments was collected using the tests provided with the OSS. Most of the tests included in the OSS are those whose results are passed. In this study, since the coverage with Pass and Fail test results is needed, we manually seeded a part of the code to create a program including bugs. The numbers in the tables in this section are rounded up at the fourth decimal point.

3.1 Target Program

We choose the following two programs as target programs for the experiments:

- **scikit-learn** is a library for machine learning. A bug is seeded in the program "kernel_approximation". The LOC of the source code is 1011. Tests attached to the library collect the coverage. The tests are given additional random seed values to increase the number of tests. A total of 288 coverage tests were generated, including 138 failed tests and 150 successful tests, and are used in our experiments.

- **Html5lib** is a library for parsing HTML. A bug is seeded in the program "serializer". The LOC of the source code is 409. Tests attached to the library collect the coverage. A total of 235 coverage tests were generated, including 23 failed tests and 212 successful tests, and are used in our experiments.

The proposed method is applied to the coverage collected from the above two OSS. Compare three bug suspiciousness measurement methods including the proposed method. The measurements compared are the following:

- proposed-measure : A difference between outputs of original coverage that fails the test and the virtual coverage.

- one-hot : An output of coverage which only a certain line is executed.
- Ochiai : Suspiciousness based on Ochiai described in Section 1.

3.2 Experiment 1

The score of the suspiciousness and the rank for seeded buggy line 195 are calculated by three measurement methods. Table 2 shows the result of the methods.

Table 2: Suspiciousness and Rank of The Buggy Line

measurement	suspiciousness	rank
proposed-measure	0.615	1
one-hot	0.525	21
Ochiai	0.818	1

Table 3: Top 5 of proposed-measure in experiment 1

Rank	Line	suspiciousness
1	195	0.615
2	158	0.589
3	199	0.007
4	194	0.006
5	981	0.005

Table 4: Top 5 of one-hot in experiment 1

Rank	Line	suspiciousness
1	720	0.528
2	981	0.528
3	668	0.527
4	449	0.527
5	951	0.527

Table 5: Top 5 of Ochiai in experiment 1

Rank	Line	suspiciousness
1	195	0.818
2	158	0.818
3	186	0.716
4	24	0.716
5	14	0.716

The proposed-measure and Ochiai ranked the line seeded the bug first, while one-hot ranked 21st.

Tables 3, 4 and 5 display the top five in the ranking based on suspiciousness score of proposed-measure, one-hot and Ochiai respectively. The line 195 is the line seeded a bug.

Table 3 shows that the measurement classifies properly the buggy line and other lines. Line 158's score closes to line 195's score. This is because their appearances in the coverage are the same. In contrast, from Tables 2 and 4, we can see that the buggy line is ranked in 21th. There is actually not much of a difference between the suspiciousness score of the buggy line and the first ranked line. As the result in Table 5, the buggy line 195 and non-buggy line 158 are tied for the top.

3.3 Experiment 2

The score of the suspiciousness and the rank for seeded buggy line 330 are calculated by three measurement methods. Table 6 shows the result of the methods.

Table 6: Suspiciousness and Rank of The Buggy Line

measurement	suspiciousness	rank
proposed-measure	0.715	1
one-hot	0.526	1
Ochiai	1.0	1

Table 7: Top 5 of proposed-measure in experiment 2

Rank	Line	suspiciousness
1	330	0.715
2	331	0.019
3	329	0.011
4	325	0.007
5	327	0.007

Table 8: Top 5 of one-hot in experiment 2

Rank	Line	suspiciousness
1	330	0.526
2	327	0.518
3	337	0.514
4	321	0.512
5	305	0.512

Table 9: Top 5 of Ochiai in experiment 2

Rank	Line	suspiciousness
1	330	1.0
2	329	0.804
3	336	0.793
4	335	0.793
5	325	0.793

All of the measurements ranked the line seeded the bug first.

Table 7 shows that the measurement distinguished the buggy line and the others. In Tables 6 and 8, the buggy line is ranked in the first. However, similarly to the experiment for scikit-learn, the suspiciousness score of 21st is 0.499. This score hardly

differs from the score of the first. As the result in Table 9, the buggy line 330 got score 1.0. Because this line is executed in all failed tests.

4 DISCUSSION

In both experiments, there is not much difference among the top 20 of the score that one-hot calculated. On the other hand, In both cases, we can see that the proposed-measure properly scored the buggy line. There is a large difference between the Top score and other lines. It is examined using Gini coefficient whether the three measurements give properly higher scores to buggy lines, and lower scores to non-buggy lines. Since some scores of proposed-measure contain negative values, the difference values from the minimum are used in the calculation. Tables 10 and 11 show Gini coefficients of the three methods in the two experiments. Each of the coefficients in the Table indicates that one-hot gives one-hot gives a fair score to each line, whereas the other two measurements give an unbalanced score to each line. From these results, we believe that the measurement proposed-measure is more effective than one-hot in whether a buggy line is determined a high suspiciousness score.

Table 10: Gini Coefficients in Experiment 1

proposed-measure	one-hot	Ochiai
0.637	0.079	0.323

Table 11: Gini Coefficients in Experiment 2

proposed-measure	one-hot	Ochiai
0.783	0.028	0.973

From the results of the experiments, It is confirmed that the proposed method is effective to localize faults. The effectiveness of the proposed method can be concluded it is comparable to Ochiai and is more effective than the one-hot method used in previous studies.

In the scikit-learn experiment, line 158 ranked in second place had the same appearance in coverage as line 195 containing the true bug, that is, if one is executed, the other is also executed. Therefore, it is reckoned from the coverage that both lines are equally likely to cause a bug. This is similar to SBFL without DNN, which calculates the probability of bugs based on the coverage, and thus cannot distinguish between lines that have the identical occurrence in the coverage. This results in a problem that a impact on bugs is determined equally. This problem can be solved by satisfying branch coverage to augment unique coverage in some cases, but this is not the case for unbranched programs. Therefore, it is necessary to narrow down the lines that have a close probability of each other by using information extracted from execution traces, source codes, error messages, etc.

In both experiments, the method was applied to more than 200 test cases and dozens of failed tests. The experiments are built on adequate failed tests and passed tests. In repositories that are a collection of actual bugs, it is frequent that there are only one or two failed tests. In order to ensure the effectiveness and generality of the method, it is required to apply the method to such actual bugs.

5 CONCLUSION

This paper proposes a method that combines SBFL and DNN. The proposed method trains a network that predicts test results based on coverage of a source code, and measures the impact of each line on bugs from the trained network, and provides a ranking based on the measurement. In measuring the suspiciousness of bugs, we focus on the difference between the output of the original coverage and the output of the virtual coverage that altered the original coverage, and treat the difference value as the impact of the bug. In the experiment, we obtained better results than the conventional measurement method, which treats the output of a virtual coverage in which only a certain line is executed as suspiciousness, and the proposed method is comparable to Ochiai.

In future work, we want to apply this method to SBFL in code blocks instead of lines. The occurrence of lines in a block must be the same. Hence, if a line is determined to be highly possibly buggy, any line in the block containing that line may be the same possibly buggy. In addition, the experiments in this paper applied the method against artificially seeded faults in the projects that have adequate quantity of test cases to evaluate the effectiveness. We would like to conduct experiments against actual bugs such as Defects4J[12] to further evaluate the effectiveness of the proposed method.

ACKNOWLEDGEMENT

Part of this work is supported by fund from Mitsubishi Electric Corp.

The research is also being partially conducted as Grant-in-Aid for Scientific Research A (18H04094) and C (21K11826).

REFERENCES

[1] Claire Le Goues, Michael Dewey-Vogt, Stephanie Forrest, and Westley Weimer, “A Systematic Study of Automated Program Repair: Fixing 55 out of 105 Bugs for $8 Each,” ICSE, pp.3-13 (2012)

[2] Jaechang Nam, Song Wang, Yuan Xi, and Lin Tan, “A bug finder refined by a large set of open-source projects,” Information and Software Technology, Vol.112, pp.164–175 (2019)

[3] Sunghun Kim, Thomas Zimmermann, Kai Pan, and E. James Jr. Whitehead, “Automatic Identification of Bug-Introducing Changes,” 21st IEEE/ACM International Con-

ference on Automated Software Engineering (ASE'06), pp.81-90 (2006)
[4] Sokratis Tsakiltsidis, Andriy Miranskyy, and Elie Mazzawi, "Towards Automated Performance Bug Identification in Python," 2016 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), pp.132-139 (2016)
[5] Keigo Matsushita, Masaki Matsumoto, Kazuhiko Ohno, Takahiro Sasaki, Toshio Kondo, and Hiroshi Nakashima, "A Debugging Method Based on Comparison of Execution Trace," Symposium on Advanced Computing Systems and Infrastructures (SACSIS), Vol.2011, pp.152-159 (2011) (in Japanese)
[6] Rui Abreu, Peter Zoeteweij, and Arjan J. C. van Gemund, "On the accuracy of spectrum-based fault localization," Testing: Academic and Industrial Conference Practice and Research Techniques, pp.89-98 (2007)
[7] W. Eric Wong, Vidroha Debroy, Yihao Li, and Ruizhi Gao, "The DStar method for effective software fault localization," IEEE Transactions on Reliability, Vol.63, No.1, pp.290-308 (2014)
[8] Takuma Ikeda, Kozo Okano, Shinpei Ogata, and Shin Nakajima, "Localization of Fault Methods and Ablation of Execution Traces Using A Machine Learning Model to Classify Test Results," IEICE Technical Report, Vol.121, No.416, pp.13-18 (2022) (in Japanese)
[9] Zhuo Zhang, Yan Lei, Xiaoguang Mao, Meng Yan, Ling Xu, and Xiaohong Zhang, "A study of effectiveness of deep learning in locating real faults," Information and Software Technology, Vol.131, No.1, pp.1–16 (2021)
[10] W. Eric Wong, Vidroha Debroy, Richard Golden, Xiaofeng Xu, and Bhavani Thuraisingham, "Effective Software Fault Localization Using an RBF Neural Network," in IEEE Transactions on Reliability, Vol.61, No.1, pp.149-169 (2012)
[11] "scikit-learn: machine learning in Python — scikit-learn 1.1.1 documentation," https://scikit-learn.org/stable/ (referred April 27, 2022)
[12] René Just, Darioush Jalali, and Michael D. Ernst, "Defects4J: A database of existing faults to enable controlled testing studies for java programs," in Proceedings of the 2014 International Symposium on Software Testing and Analysis, pp.437–440 (2014)

Automated Scoring Gateball Realtime with Camera, YOLO, and RTSP

Trong Pham Dinh*, Mikiko Sode Tanaka** , Ami Uchikata*, and Koki Ishioka*
* Kanazawa Institute of Technology, Japan
** International College of Technology, Japan
* {c1295108,c1295153,c1295212}@planet.kanazawa-it.ac.jp
* sode@neptune.kanazawa-it.ac.jp

Abstract - Gateball's popularity has gradually decreased over the past 20 years. This research aims to build a system that allows players can enjoy playing with basic rules without referees such as bowling. Because of it, the player can create new playstyles based on it. By reducing the difficulty for beginners, we can increase the coverage of the Gateball. In this research, we set the camera on the gates that were built with Raspberry Pi and the Camera module to collect data when the ball passes through the gate, then transfer it to the PC in real-time with Real-Time Streaming Protocol. In the PC, it uses the YOLO model to detect the ball and then track it. The system worked well in a wireless network with the latency of transferring data between these devices being under 500ms. This system was able to check whether the ball pass through the gate or not and calculated the score for each ball passing automatically. However, with the speed of the ball increasing, the accuracy is significantly reduced.

Keywords: Gateball, Realtime Object Detection, RTSP, IoT

1. INTRODUCTION

Gateball is a sport and strategy game that was invented by Eiji Suzuki of Memuro town, Hokkaido in 1947 and gradually became popular throughout Japan from the 1960s to the 70s and then following that popularity Gateball began spreading from Japan to the world, starting from Asia countries to the United States during the 1980s [1]. However, in the last 20 years, the number of Gateball's players is rapidly decreasing from 567,232 people in 1996 to 94,703 people [2]. The reason for this decrease is the average age of players and the difficulties of Gateball rules [3].

96% of the 90,000 members of the Japan Gateball Federation are over 60 years old. Therefore, Gateball is well-known as a sport for elder people. There are many problems with this high average age. One of them is the physical condition. In every sport, the player has required to be in the best physical condition during the match to be able to play whole the match and make the right judgment at every moment. This is hard for elder people because the human body naturally is increasingly weaker and the brain's reaction speed is slower than when they were young. Therefore, the possibility that they cannot participate in the match or must take a break in the middle is extremely high. If there is a person who can't participate they are forced to stop the match. Moreover, with old age, the time they belong on the team will definitely be shorter when compared with young people because of the heath problem and transport problems. Therefore, it is important to look for more young people to join this sport.

In addition to the age problem, the complexity of the rules and the gameplay of gateball is also an obstacle for the new beginner. As shown in Figure 1, Gateball is a game played by two teams, each with five players that leading team uses red balls (odd numbers), and the following team uses white balls (even numbers). The player strikes the ball from the start area towards the first gate to pass the ball through it as shown in Figure 1.

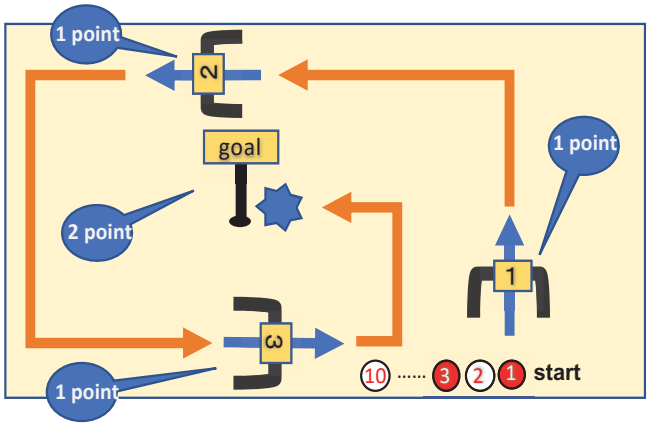


Figure 1: The rule of Gateball

The player aims to pass his/her ball through the second gate and the third gate in that order. One point for every time the ball passes through the first, second, and third gates, while two points are awarded when the ball hits the goal. The team, therefore, scores 25 points if all five team members finish. Because of this, the player has to think strategically about how to get their team's ball to hits the goal as much as possible and kick the opposing team's ball out. In order to come up with the most optimal strategy, players must understand the rules of the game, their team's strengths, and weaknesses as well as predict the opponent's playstyle. This will take a lot of time in order to do, which the beginners don't have. We need to find a way to keep them playing, experienced enough to be able to come up with strategies based on that.

From the above, in this research, we experimented with applying IoT devices and object detection technology to automate scoring in Gateball. Then, based on this system we suggest new and simple gameplay that is more friendly for beginners to be able to start playing this sport.

2. RELATED WORK

2.1 Detection of Animals Combining RGB and Depth Images

In the field of animal behavior and ecology study, observing the movement of animals is the most optimal research method. However, if the researcher does this his/herself there will be many problems such as it requires a lot of time and effort or damage to the animal's natural habitat during observation. To solve this problem, we can use laser sensors and a hybrid camera to track animals automatically [4]. In this research, he used the hybrid camera to catch any movement of animals and redraw the movement of the observed animal. Moreover, in order to avoid noise and error in image recognition, he used depth images and foreground extraction using the RANSAC algorithm. As a result of this experiment, it was able to recognize the movement of the target animal and track it correctly.

2.2 Support System for Athletes and Coaches Using Video-Based

In order to improve the efficiency of the player's training, technologies such as IoT (Internet of Things) and AI (Artificial Intelligence) are increasingly being applied widely in the training process of the players. An example of them is the video feedback system[5] which provides support for conducting specialized training and maintaining discipline. In this research, he set up video cameras in several places in order to record the process of training, the coach will be able to check the movement of athletes from various angles. However, while recording the video the camera will be not able to take pictures. To solve this problem, he decided to build a mobile video system using P2P technology.

3. AUTOMATED SCORING SYSTEM

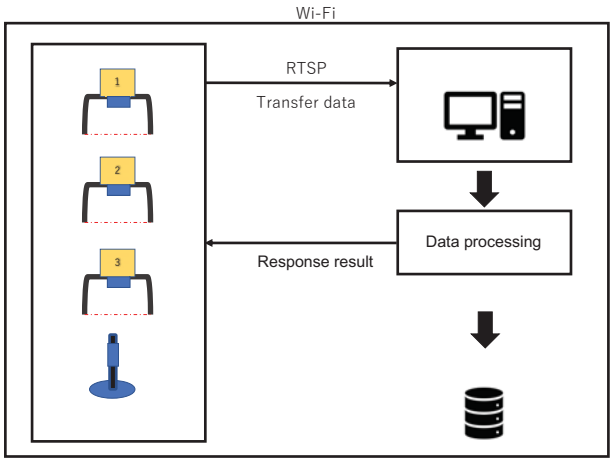


Figure 2: Automated Scoring System

This system calculates the score of each player based on image data that were collected from the camera module above the gate and sends the response back to the camera module after finishing calculating the score. This was made up of 2 parts as the Figure2. The first part is a camera module which is responsible for collecting data during the match and sending it to the local PC. The second part is a data processing machine that receives data from remote camera modules, then processes them using object recognition algorithms in order to calculate the score. Currently, in order to play Gateball, we need three referees who are a referee, an assistant referee, and a recorder. With this system, we can automate and calculate the score of each player without referees with less number of required people in the match and gain more 3 available players thereby minimizing the problem of lack of players.

3.1 CAMERA MODULE

In this section, we will explain the camera module which is part of collecting data during the match and sending it to the local machine in order to calculate the score of each player. This camera module is made up of three parts:

(1) Raspberry Pi Zero W
(2) Camera
(3) Battery

By using Raspberry Pi Zero W, it allows this camera module to be able to use the RTSP protocol which is used to transfer video or image data in real-time with a latency of less than 500ms. We use the removable camera in this module because there was not a Raspberry Pi Zero W that has a built-in camera and by using the removable camera, we will have more options to test with many cameras thereby deciding which ones are the most optimal for the system. Currently, we are using a 10000mAh battery in this module. However, we realized that it is too much for this small module and are planning to create a new design of the module with a small 500mAh battery in order to make it become easier to set in the gate.

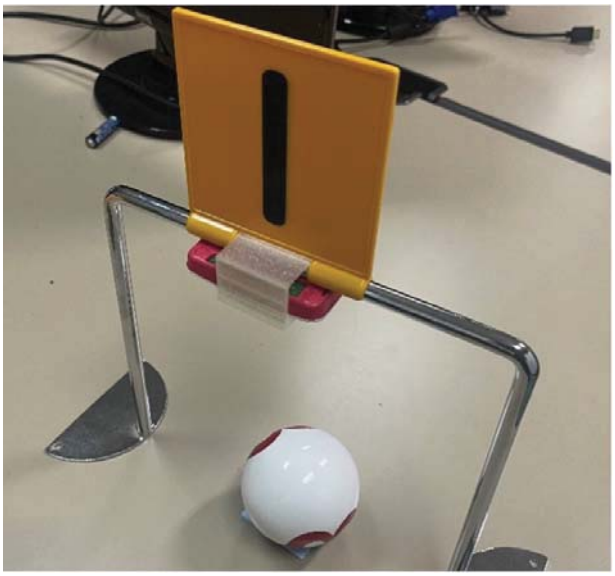


Figure 3: Camera Module

3.2 SCORE CALCULATE

This is the most important part of the automated scoring system, we use Jetson Xavier NX as the local machine

which is used to receive the image data sent from remoted camera modules with RTSP protocol and process them with the YOLO object detection algorithm in order to calculate the score of the player. YOLO(You Only Lock Once) is one of the most popular object detection algorithms recently, it uses neural networks to provide real-time object detection with extremely fast and accurate. In mAP measured at 0.5 IOU YOLOv3 is equal to Focal Loss but faster about 4 times. In this research, we decide using this algorithm to detect balls from the collected data because it allows real-time detection, light, high precision, and easy to train the model. In this part, we use python3 which is a high-level, interpreted, and general-purpose programming language to work with the YOLOv3 model. YOLOv3 was created based on C/C++ language which was the language python was built on. Because of it, the system can work smoothly without conflicts such as syntax or compiler errors. In this local machine, the image data with the RTSP protocol was covert to a 3-dimension vector [B][G][R] in the range 0 ~ 255 and then using the YOLOv3 model to check every pixel to detect the bounding box of the object in the image. We use the detected bounding box to calculate the coordinates of the ball to check whether the ball passes through the gate or not. YOLO model not only responds to the bounding box but also responds to the class of detected objects. With the response class, it is able to know which ball passed through the gate and calculate the score of the player. To summarize this part, the calculated score machine work as shown in Figure 4.

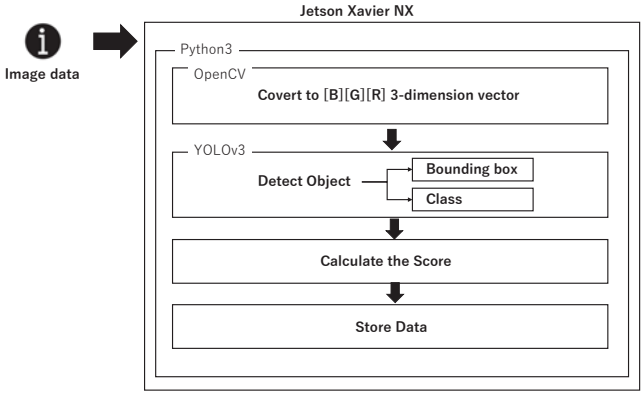


Figure 4: Score Calculate Process

4. EXPERIMENTAL RESULT

4.1 DATA SET

The dataset used in this experiment is made by our members which includes 1000 sample pictures that were labeled in YOLO format as shown in the table1 .

Table 1. The YOLO format used for dataset

Class	X	Y	Width	Height
0~10	0~1	0~1	0 ~	0 ~

4.2 RESULT

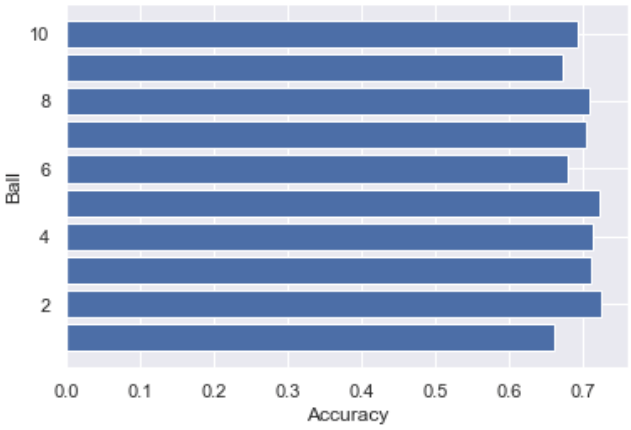


Figure 5: The accuracy of each ball

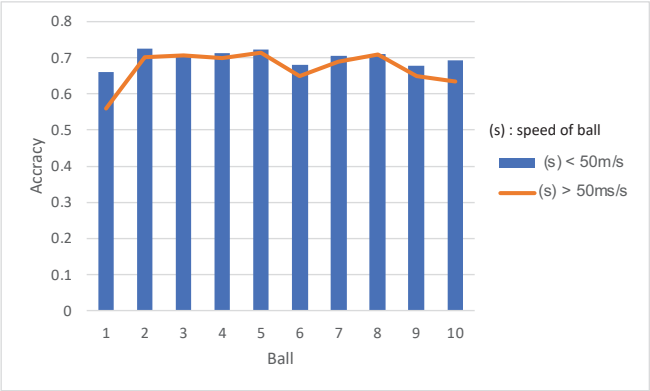


Figure 6: The accuracy and the speed of the ball

In this experiment, we test the accuracy with our custom YOLO model which was trained with 700 pictures, 70 pictures for each class. First, we test with random 300 testing pictures which split from the dataset. The result of this first test shows that with the image the accuracy for all classes is above 0.6 as Figure 5. The highest accuracy is the ball 2 and the lowest is ball 1. From this first experiment result, we understand that for the balls with the same characteristics, the accuracy of detection is lower than the others. For example the ball 1 and the ball 10 that has a part show number "1" so when the image of the ball only shows "1" the model can't detect it well.

In the second experiment, we test the system with the raw video that was transferred from the camera module on the gate. We test with 2 methods in this experiment. The first one is with the slow speed of the ball being below 50ms/s. The second one is with a high speed that is above 50ms/s. In Figure 6, It shows us that with the high speed of the ball the accuracy of detection is decreasing. The reason is when the ball passed the gate at high speed the camera had difficulty capturing clearly the shape of the ball, when the difference between the input image and the learned image is large it is difficult for the system to recognize it correctly.

5. NEW GAMEPLAY

From the experimental result, the system shows that it is able to work with real games in Gateball in a simple mode like calculating score whenever the ball passed through the

gate. We suggest two gameplays based on this system which is easier for beginners who do not know a lot about the rules.

(1) The players will start their turn according to the number of balls from the start area. For each time the ball passes through the gate the player will gain one point. Who has got 10 points first will be the winner.
(2) The player will be divided into two teams based on the number of the ball (even-odd). One team will be the attacker, the other will be the defender. For one turn, the player can strike the ball once. With every ball that passes through the gate his/her team will get one point. The defender team will think about how to block the attacker pass the ball through the gate. If the attacker team gets 20 point they will be the winner. However, if the attacker team hasn't gotten the 20 points after 25 turns the defenders' team will be the winner.

6. SUMMARY

In this research, we build a system named Automated Scoring system using IoT and AI in order to solve the high average age and lack of player problem in Gateball. The result of the experiment shows that the system is able to work on simple gameplay.

ACKNOWLEDGMENTS

This research is supported by I-O DATA Foundation.

REFERENCES

[1] World Gateball Union: History of Gateball https://gateball.or.jp/wgu/play/play_04.html
[2] The decline in Gateball's population and reason https://www.news-postseven.com/archives/20180307_657061.html?DETAIL, Access 2022.6.8
[3] Yazaki Wataru, 0230305 Study on Sports Activities of the Elderly: Declining Gateball Population, Annual Meeting of the Japanese Society of Physical Education, 1992, Volume 43A, 43rd (1992), p. 167-, Published 2017/08/25, Online ISSN 2433-0183, https://doi.org/10.20693/jspeconf.43A.0_167, https://www.jstage.jst.go.jp/article/jspeconf/43A/0/43A_167/_article/-char/ja
[4] Tsune Nada, McKinness James, Yasuo Nagai, Asami Tsuchida, Koji Masuda, Detection of Animals Combining RGB and Depth Images, Proceedings of the Fuzzy System Symposium of the Japan Intelligent Information Fuzzy Society, 2013, Vol. 29, 29th Fuzzy System Symposium, Session ID ME1-1, p. 48, Release date 2015/01/24, https://doi.org/10.14864/fss.29.0_48, https://www.jstage.jst.go.jp/article/fss/29/0/29_48/_article/-char/ja
[5] Tomokazu Miura, Support System for Athletes and Coaches Using Video-Based, Journal of Educational Systems and Information Systems, 2020, Vol. 37, No. 3, p. 185-191, Published 2020/07/01, Online ISSN 2188-0980, Print ISSN 1341-4135,https://doi.org/10.14926/jsise.37.185, https://www.jstage.jst.go.jp/article/jsise/37/3/37_370304/_article/-char/ja

Automatic Derivation of a Transition Model from a Japanese Requirement Specification under a Restricted Grammar

Koki Shimokawa†, Hiroya Ii†, Maiko Onishi‡, Shinpei Ogata*, and Kozo Okano*

†Graduate School of Science and Technology, Shinshu University, Japan
‡Graduate School of Humanities and Sciences, Ochanomizu University, Japan
*Faculty of Engineering, Shinshu University, Japan
* {ogata,okano}@cs.shinshu-u.ac.jp

Abstract - In this paper, we propose a method for extracting state transitions, which are necessary elements for model checking, from requirement specifications written in a natural language. Requirement specifications are often written in a natural language and may contain ambiguous expressions. Model checking is a method to check for such ambiguous expressions using mathematical techniques. The proposed method extracts state transitions from requirement specifications based on restricted grammars. We also present the results of an experiment to evaluate the effectiveness of the proposed method.

Keywords: Natural Language Processing, Requirement Specification, Model Checking.

1 INTRODUCTION

Requirement Specification documents used in software development are often written in natural languages [1]. Natural languages may contain ambiguous expressions in the text. This leads to the discovery of bugs in the testing process, which results in significant rework costs. Model checking approaches are regarded as a useful method to verify whether the specification is good requirement or not. Model checking, however, requires specialized knowledge such as temporal logic. Thus, we believe that a method to automatically derives models from requirement specifications in natural languages will help designers and developers to share models without contradiction.

In our previous study, we proposed a method for automatically extracting state variables and state transitions using morphological and syntactic analysis for the purpose of automatically creating state transition diagrams [2]. The results show that the recalls of state variables and variable values are 1.00 for 36 sentences, and the precisions of state variables and variable values are 0.93 and 0.83. However, the significant increase in parsing rules resulted in a major problem of extracting unnecessary elements.

In this article, we consider that restricting the input syntax to some extent can solve the above-mentioned. Section 2 explains the terminology. Then Section 3 describes previous studies. Section 4 discusses related research. Section 5 gives the proposed sentence structure. Section 6 describes the experimental evaluation methodology. Section 7 describes the results. Section 8 discusses on the results. Finally, Section 9 provides conclusion and future outlook.

2 PRELIMINARIES

2.1 Morphological Analysis

Morphological analysis is the process of dividing input sentences into words and recognizing their parts of speech and conjugations. In this paper, our research group use an existing tool, kuromoji [3]. It is an open-source Japanese morphological analyzer developed in Java, and is capable of word segmentation, part-of-speech tagging.

2.2 XPath

XPath stands for XML Path Language [4]. It is a concise syntax for extracting specific parts of a document in XML format. Using this tool, you can follow the elements you want to extract from a tree of sentences. This tool allows you to follow the elements you want to extract from a tree-structured document like a file path.

2.3 NuSMV

NuSMV is a software tool for the formal verification of finite state systems. Other model checking tools besides NuSMV include SPIN and UPPAAL. It has been developed jointly by FBK-IRST and by Carnegie Mellon University [5]. NuSMV allows to check finite state systems against specifications in the temporal logics CTL and LTL. The input language of NuSMV is designed to allow the description of finite state systems that range from completely synchronous to completely asynchronous. The NuSMV language provides for modular hierarchical descriptions and for the definition of reusable components. The basic purpose of the NuSMV language is to describe the transition relation of a finite Kripke structure [6]. This provides a great deal of flexibility, however, it can introduce danger of inconsistency for non expert users at the same time.

3 PREVIOUS RESEARCH

In order to extract the elements of the transition model, we decomposed sentences into words and tagged them with parts of speech using the morphological analyzer kuromoji. Further, our research group created a syntactic tree using the CYK method [7]. Then, the proposed method classified the syntax with XPath and extracted the specified elements. As a result, the recalls of state variables and variable values were 1.00, and the precisions were 0.93 and 0.83, showing high accuracy for the 36 input sentences described in the description of the Electric Pot (7th edition) [8]. However, there is a wide variety of word orders in the input sentences, such as inverted sentences and emphatic sentences. Our research group considered that creating rules for syntax tree creation and element extraction in consideration of such cases would require an unrealistic number of rules. In fact, the number of syntax tree creation rules for 36 input sentences was 163, and the number of element extraction rules increased as well, resulting in the extraction of extra elements and reducing the precision.

4 RELATED WORK

There are a number of studies that convert specifications written in natural language to UML (Unified Modeling Language) or Model Checking.

In order to extract the components of the model, there is an attempt to use a dependency parsing instead of manually creating phrase structure rules [9]. They used "ChaboCha" [10] for the dependency parsing. "ChaboCha" uses machine learning called Support Vector Machine, so there is no need to set rules for parsing [11]. As a result, all state variables and variable values could be extracted as in the extraction with clause structure rules, but extra elements were extracted.

In addition, there are also studies that have attempted to generate class diagrams by extracting elements from English sentence specification based on the SBVR (Semantics of Business Vocabulary and Business Rules) [12] standard established by OMG (Object Management Group) [13]. In order to map SBVR to the UML class model, it is a matter of extracting SBVR vocabularies from SBVR rules and translating SBVR vocabularies into the basic elements of the UML class model (class names, relations, etc.). The SBVR vocabulary consists of definitions of all specific terms and concepts used in the course of business. They classify SBVR vocabularies as common nouns, proper nouns, verb concepts, quantification("a", "an" and prefixed with "s"), partitive fact types ("is part of," "included in," or "belong-to"), etc. using SBVR rules and natural language processing. And then, they use their own rules to map from the classified SBVR vocabulary to the elements needed for the class diagram. For example, a common noun maps to a class name. Characters or unary fact type(associated an object type) map attribute name. As a result, class diagrams could be generated with high accuracy.

There are also studies that convert from natural language processing to SBVR representation [14]. It has been reported that when a meta-model is created and requirement specifications in natural languages are converted into that meta-model, the formal logic of SBVR allows requirements to be easily processed by machines and element extraction to be performed with high accuracy. Similar study is being conducted on Japanese specifications. Study is being conducted to develop a support tool that automatically generates UML text from Japanese requirement specifications using natural language processing [15]. The tool automatically generates object diagrams, class diagrams, sequence diagrams, use case diagrams, and state machine diagrams. As a result, they were able to successfully generate UML with high accuracy. However, it has the disadvantage of being cumbersome when the proposed method does not match the method that the designer wants to adopt.

Another study proposed a method to automatically generate a temporal logic formula that can be mapped to an inspection formula as an intermediate process of extracting state-transition relations from conditional sentences that appear in specifications [16].

5 PROPOSED METHOD

The ultimate goal of this research is to automatically create state transition models from requirement specifications written in natural languages and model checking. In order to improve the aforementioned problem, we devised a method to extract elements by defining a certain restricted stylistic description in our previous study [17]. Then, a GUI element support tool was devised to present the extraction results to the specification writer in real time at the stage of creation, and to encourage the writer to correct the sentence. In this study, we decided to devise an input sentence that can be easily converted to a state transition model description automatically, referring to previous studies. The following sections describe the elements to be extracted, the input sentence creation method, the proposed grammatical structure, and the outline of the element support tool.

5.1 Extracted Elements

The model used in this study is NuSMV. The description of NuSMV consists of sections defining the set of state variables and variable values, the initial state, and the destination state [17].And we believe that if we can extract all the necessary elements for the destination state, we can extract all of them. Because there is always a set of variable

values for the state variable in the description of the destination state.

Therefore, we decided to extract the five elements of state variable, pre-transition variable value, transition condition, user behavior, and post-transition variable value in order to also enable automatic creation of state transition diagrams from requirement specifications written in natural language. The description part of the destination state, which represents the five elements of state variables (red), pre-transition variable values (blue), transition conditions (green), user behavior (purple), and post-transition variable values (yellow), is shown in Figure 1.

In addition, state variable, pre-transition variable value, user behavior, and post-transition variable value are essential to execute NuSMV.

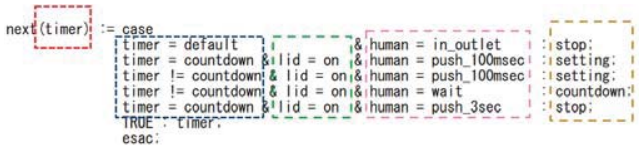


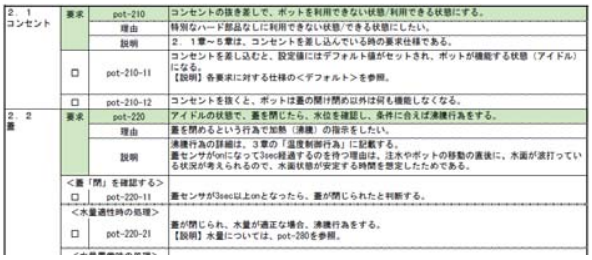
Figure 1: Example of model checking description

5.2 Sentence Preparation Procedure

We created specification sentences with restricted grammars in order to make the sentences easy to be automatically extracted into the NuSMV model. The requirements specification document used as reference in this study is the Electric Pot (7th edition) [8]. In creating the restriction grammar, we first wrote the NuSMV model description of the Electric Pot (7th edition). The reason for creating a NuSMV description is to find transition state statements that are omitted in the requirements specification description so that NuSMV can be executed. For example, "When the lid is open and the user closes it, the lid closes" is not included in the specification, so it was complemented. Two students with knowledge of NuSMV read the specification description (in Figure 2) and created a NuSMV model.

Next, they checked whether the created model satisfy timed properties expressed in LTL. The letter "U '' between variables "condition '' and "check '' in the LTLSPEC description stands for "until.''The whole expression "condition until check '' stands for "condition" holds until "check" holds. An example of condition phrase in a natural language corresponding "condition '' variable is "the user is pressing or releasing the hot water button while the water is being supplied. '' Another example of those is "the water temperature is low when the user presses the boil button. ''

On the other hand, an example of condition phrase in a natural language corresponding "check '' variable is "the safety lock does not turn on and the water does not boil ''



2．1 コンセント	要求	pot-210	コンセントの抜き差しで、ポットを利用できない状態/利用できる状態にする。
		理由	特別なハード部品なしに利用できない状態/できる状態にしたい。
		説明	2．1章～5章は、コンセントを差し込んでいる時の要求仕様である。
	□	pot-210-11	コンセントを差し込むと、設定値にはデフォルト値がセットされ、ポットが機能する状態（アイドル）になる。 【説明】各要求に対する仕様の<デフォルト>を参照。
	□	pot-210-12	コンセントを抜くと、ポットは蓋の開け閉め以外は何も機能しなくなる。
2．2 蓋	要求	pot-220	アイドルの状態で、蓋を閉じたら、水位を確認し、条件に合えば沸騰行為をする。
		理由	蓋を閉めるという行為で加熱（沸騰）の指示をしたい。
		説明	沸騰行為の詳細は、3章の「温度制御行為」に記載する。 蓋センサがonになって3sec経過するのを待つ理由は、注水やポットの移動の直後に、水面が波打っている状況が考えられるので、水面状態が安定する時間を想定したためである。
	<蓋「閉」を確認する>		
	□	pot-220-11	蓋センサが3sec以上onとなったら、蓋が閉じられたと判断する。
	<水量適性時の処理>		
	□	pot-220-21	蓋が閉じられ、水量が適正な場合、沸騰行為をする。 【説明】水量については、pot-280を参照。

Figure 2: Example of referenced specification description

(the contents of line 65). When performing model checking, the contents described in the 65th to 69th lines of Listing 1 were substituted into the "check" variable of the LTL expression on the 73rd line. As a result, the model checker returned true on all checks. Therefore, it was confirmed that no counterexamples for the five LTL properties(lines 65 to 69) were found.

In this way, it was found that the state transition model corresponding to water boiling function of the Electric pot (7th edition) satisfied properties from 1 to 5(lines 65 to 69) We believe that the model created can be used for safety verification. In addition, it is also possible to intentionally output a counterexample in model checking to check whether a specific operation is possible. For example, "If the 'condition ' are met when the lid was open, '' is model checked. As a result, as shown in Fig. 3, desired counterexample was output.

Thus, it was possible to confirm that there was no problem with the model. Since it was confirmed that there was no problem in the model, We consider a statement that includes the elements needed to create a model.

Listing 1: Example of NuSMV model

```
1  MODULE main
2  VAR
3        Controller : {supply, supply_stop,
             boil, can_use, cann_use, idle};
4        Human : {kp_s, p_b, wait, i_o, u_o,
             c_r, o_r,r_s, on_slb, off_slb,
             adj_wl};
5        Water_level : {proper,not_proper};
6        Rid : {open, close};
7        Temp : {low, high};
8        S_L:{on,off};
9
10 DEFINE
11   otherwise :=TRUE;
12 ASSIGN
13                init(Controller):= can_use;
14                init(Human) := wait;
15                init(Water_level) :=
                      not_proper;
16                init(Rid):= open;
17                init(Temp):= low;
18                init(S_L):= on;
19
20                next(Controller) := case
21                         Human = kp_s & Temp =
```

```
                        high & Controller
                        != cann_use &
                       Rid = close & S_L
                        = off : supply;
22                  Controller = supply &
                        Human = r_s :
                        supply_stop;
23                  Human = p_b &
                        Water_level =
                        proper & Temp =
                        low & Rid = close
                        : boil;
24                  Human = i_o &
                        Controller =
                        cann_use: can_use;
25                  Human = u_o &
                        Controller =
                        can_use : cann_use
                        ;
26                  Controller = can_use :
                         idle;
27                  Controller = boil :
                        can_use;
28                  Human = c_r &
                        Controller = idle
                        & Water_level =
                        proper : boil;
29                  otherwise : Controller
                        ;
30                  esac;
31
32          next(Human) := case
33                  Rid = open &
                        Water_level=
                        not_proper :
                        adj_wl;
34                  Human = kp_s : {kp_s,
                         r_s};
35                  Controller != cann_use
                         & Controller !=
                        supply: {kp_s,
                        wait,p_b, u_o, c_r
                        , o_r, on_slb,
                        off_slb} ;
36                  Controller = cann_use
                        : {i_o, o_r, c_r
                        };
37                  otherwise : Human;
38          esac;
39
40          next(Water_level) := case
41                  Human = adj_wl & Rid =
                         open &
                        Water_level =
                        not_proper: proper
                        ;
42                  otherwise :
                        Water_level;
43          esac;
44
45          next(Rid) := case
46                  Human = c_r & Rid =
                        open: close;
47                  Human = o_r & Rid =
                        close & Controller
                         != boil &
                            Controller !=
                            supply : open;
48                  otherwise : Rid;
49          esac;
50
51          next(Temp) := case
52                  Controller = boil :
                        high;
53                  Human = wait : low;
54                  otherwise : Temp;
55          esac;
56
57          next(S_L) := case
58                  Human = on_slb & S_L =
                         off & Rid =
                        close & Controller
                         != supply : on;
59                  Human = off_slb & S_L
                        = on & Rid =
                        close & Controller
                         != supply : off;
60                  Human = r_s & S_L =off
                         : on;
61                  otherwise : S_L;
62          esac;
63
64  DEFINE
65          check1 := S_L = on &
                Controller = supply;
66          check2 := Rid_sensor = close
                & Controller = supply;
67          check3 := Rid = open &
                Controller = boil;
68          check4 := Water_level =
                not_proper & Controller =
                 supply;
69          check5 := Temp = low &
                Controller = supply;
70          condition1 := (Controller =
                supply) -> ((Human = kp_s
                )|(Human = r_s));
71          condition2 := (Human = p_b)
                ->(Temp = low);
72          condition := condition1 &
                condition2;
73  LTLSPEC !(condition U check2)
```

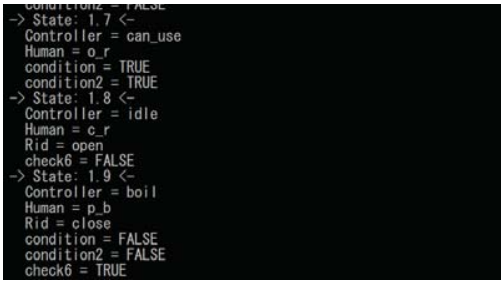


Figure 3: Counterexamples for boiling transitions

5.3 Proposed Sentence Structure

As shown in Section 5.1, the proposed sentence structure consists of five elements. Figure 4 to Figure 7 show the sentence structures and examples of sentences using the Element Extraction Support Tool.

5.3.1 Proposed Sentence (1)

The first is a sentence type that describes all elements of the state variable, pre-transition variable value, transition condition, user behavior, and post-transition variable value, as shown in Figure 4.

Figure 4: Proposed sentence structure (1)

5.3.2 Proposed Sentence (2)

The second is a sentence structure in which elements other than transition conditions are arranged as shown in Figure 5.

Figure 5: Proposed sentence structure (2)

5.3.3 Proposed Sentences (3) and (4)

The third and the fourth are sentence structures (1) and (2) without the state variable. The reason for adopting this sentences is that the state variable is not implicitly described in the specification. Therefore, we will consider providing the ability to supplement the state variable with the proposed tool (in Fig. 6).

Figure 6: Proposed sentence structures (3) and (4)

5.3.4 Proposed Sentence (5)

As shown in Figure 7, the fifth is a sentence structure consisting of a pre-transition variable value and a post-transition variable. In the requirements document used as a reference [8], there are different variable names with the same meaning.

Therefore, we adopted this sentence structure in order to recognize that the variable names have the same meaning by describing this sentence.

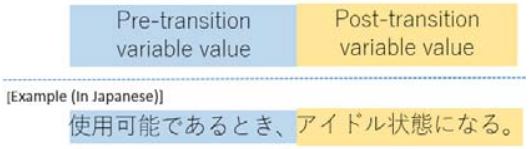


Figure 7: Proposed sentence structure (5)

5.4 Element Extraction Support Tool

With reference to the tools devised in related studies [17], we developed a tool. Figure 8 shows the overview of the tool. By analyzing input sentences written in an Excel file, the contents of the xml file describing the extracted elements are output.

Figure 9 shows the window view of the tool. The left side shows the input sentences of the excel file and the right side shows the extraction results. The tool currently provides four functions: saving of input text, deleting input text, generating a new input text, and analyzing the results of the input text.

We will discuss each of these functions in detail in the following sections.

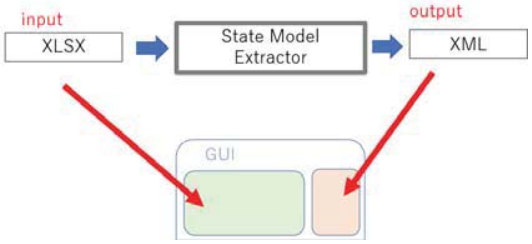


Figure 8: Tool Overview

5.4.1 Editing and Saving

The area for describing requirement specifications is a text field that can be freely edited on the tool. You can freely edit the text on the tool.

Figure 9: Developped Element Support Tools

When we press the “save” button at the bottom of the window on the left side of Figure 9 , the text file that executes the program saves the input sentences.

5.4.2 Deletion

If there is the unwanted sentence as shown in Figure 10, press the “Delete” button to the left of the component name. Then, we can delete the sentence by pressing our “Update” or “Delete” button.

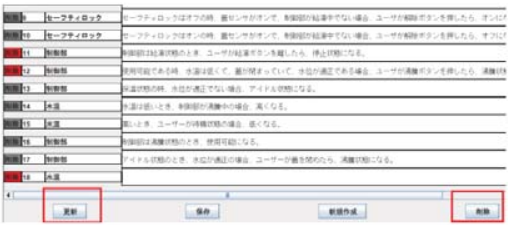


Figure 10: Delete Function

5.4.3 Generation

Pressing the “New” button at the bottom of the left window in Figure 9, the window shown in the red frame in Figure 11 appears. We then select the name of the new input statement component we are adding. If component name is not in the selection, we can type it directly in the text field.

5.4.4 Analysing Results

When we press the “Update” button at the bottom of the window, we will extract the elements from the input sentences and display them as shown in Figure 12.

6 EVALUATION EXPERIENCE

The descriptions in Chapters 2 and 3 of the Electric Pot (7th edition) [8] are translated into the NuSMV model.

Figure 11: Generating Function

Figure 12: Analysis Result Display

We then converted again into the proposed sentence structure in Japanese. As shown in Figure 13, 31 sentences are used as input sentences. The following research question is formulated to verify the validity of the proposed sentence structure.

The RQ1 is verified by calculating the reproduction recall and the precision.

RQ1: Can the proposed sentence structure be used to extract elements accurately?

RQ2: Can the number of parse tree creation rules be reduced compared to the 163 rules in the previous study [2]?

Number	Sentence
1	水位は適正でないとき、蓋が開いている場合、ユーザが調整すると、適正になる。
2	水位は適正のとき蓋センサがオンで、ユーザが給水線を超える水を入れたら、異常状態になる。
3	蓋がオンの時、ユーザが蓋を開けたら、オフになる
4	蓋がオフの時、ユーザが蓋を閉めたら、オンになる
5	制御部は使用不可能であるとき、ユーザがプラグを挿したら、使用可能になる。
6	制御部は使用可能であるとき、ユーザがプラグを抜いたら、使用不可能になる。
7	制御部は使用可能であるとき水温が高くて、ポットが使用不可能でなく、蓋が閉まっていて、セーフティロックが解除されている場合、ユーザが給油ボタンを押し続けたら給湯状態にする。
8	給湯状態のとき、ユーザが給湯ボタンを離したら、停止状態になる。
9	使用可能である時、水温は低くて、蓋が閉まっていて、水位が適正である場合、ユーザが給油ボタンを押したら、沸騰状態になる。
10	アイドル状態のとき、水位が適正である場合、ユーザが蓋を閉めたら、沸騰状態になる。
11	制御部は給湯状態のとき、電源がオンで、ユーザが給水線を超える水を入れたら、エラー状態になる。
12	給湯状態のとき、蓋がオンで、電源がオンで、セーフティロックがオンで、水位が適正である場合、ユーザが給湯ボタンを押したら、保温状態になる。
13	制御部は保温状態のとき、ユーザが蓋を閉めたら、アイドル状態になる。
14	保温状態の時、ユーザが水を入れたら、アイドル状態になる。
15	制御部は沸騰状態のとき、ユーザーが蓋を閉めたら、アイドル状態になる。
16	沸騰状態の時、ユーザが水を入れたら、アイドル状態になる。
17	使用可能状態であるとき、アイドル状態になる。
18	セーフティロックはオフの時、蓋がオンで、制御部が給湯中でない場合、ユーザが解除ボタンを押したら、オンになる。
19	セーフティロックはオンの時、蓋がオンで、制御部が給湯中でない場合、ユーザが解除ボタンを押したら、オフになる。
20	モードは高温モードのとき、蓋センサがオンである場合、ーザが保温設定ボタンを 100msec 以上押すと、節約モードになる。
21	節約モードのときに蓋が on である場合、ユーザが保温設定ボタンを 100msec 以上押すと、ミルクモードになる。
22	ミルクモードのとき、蓋が on である場合、ユーザが保温設定ボタンを 100msec 以上押すと、高温モードになる。
23	電源はオフの時、ユーザがコンセントを繋げたら、オンになる。
24	タイマは起動中のとき、蓋がオンである場合、ユーザがタイマボタンを 100msec 以上押すと、セット中になる。
25	タイマは起動中以外のとき、蓋がオンならば、ユーザがタイマボタンを 100msec 以上押すと、セット中になる。
26	タイマは起動中以外の時、蓋がオンである場合、ユーザがタイマボタンを押すのを止めて 1sec 後経過すると、カウントダウン状態になる。
27	カウントダウン状態の時、蓋がオンである場合、ユーザがタイマボタンを 3sec 以上押すと、停止状態にする。
28	ブザーはオフの時、ユーザがタイマボタンを 100msec 以上押すと、オンになる。
29	ブザーはオフのとき、ユーザがタイマボタンを 3sec 以上押すと、オンになる。
30	オンのとき、オフにする。
31	満水センサはオフの時、ユーザが給水線を超える水を入れたら、オンになる。

Figure 13: Input Sentences

7 EXPERIMENTAL RESULTS

Table 1 summarizes the results of the expreriments. It is judged that the elements are consistent even when particles or verbs are attached to the elements to be extracted, since the element specification is done in phrases. RQ1 and RQ2 can be answered as follows.

RQ1: The recall and precision for all elements are 1.00, as shown in Table 3. Therefore, it is possible to extract the elements accurately.

RQ2: The number of parse tree creation rules used in the evaluation experiment was 56, and the proposed sentence structure can significantly reduce the number of parse tree creation rules.

Table 1: Recall Precision

Elements	Recall	Precision
State variable	1.00	1.00
Pre-transition variable value	1.00	1.00
Transition condition	1.00	1.00
User behavior	1.00	1.00
Post-transition variable value	1.00	1.00

8 CONSIDERATION

We found that the proposed sentence structure greatly improves the extraction accuracy. The proposed method of sentence structure construction can be applied to a variety of sentence styles with a small number of syntax tree creation rules. In the evaluation experiment, element extraction was performed on sentences created using a sentence structure that is easy to extract elements from the NuSMV state transition model. Therefore, it is easy to imagine that the reproduction recall and precision are both 1.00 for all the elements.

In this paper, we created input sentences based on the state transition model after creating the state transition model. However, we believe that other requirement specification descriptions with the proposed sentence structure can be used to automatically create descriptions of transition relationships sentences.

However, limiting the number of input sentences has created the problem of whether users can write verbose sentences or not.

9 CONCLUSION

The goal of this paper is to create state transition models directly from requirement specifications. Therefore, we classified the elements necessary to compose a description of the transition relationship into five categories. These five elements are the elements that compose the description of the transition relationship between the variable values possessed by the state variables. We also proposed a five sentences pattern to facilitate the extraction of these elements. Moreover, the descriptions of the transition relations in the state transition model were converted into Japanese sentences with the proposed sentence structure. Then, we conducted an extraction experiment of the converted Japanese sentences using natural language processing. The result was that the recalls and precisions for all elements were 1.00. In addition, we also developed an element extraction support tool.

There are two things we plan to do in the future. The first is to define the sentence structure of the statement, including the description of the initial state and state rules. Therefore, we redefine the sentence structure of initial and transition states so that they do not cover each other. Then, we believe that all the elements necessary to create a state transition model for NuSMV can be extracted from natural languages by modifying the syntax tree creation rules and element extraction rules.

The second is to implement a guidance function in the tool that would allow the proposed sentence structure to be described. Current element support tools cannot extract elements without understanding the sentence structure proposed by the writer. Therefore, we believe that we can improve the current tool by adding a function that prompts the user to rewrite the sentence if the writer does not write the sentence with the proposed sentence structures.

Acknowledgement

This research is being partially conducted as Grant-in-Aid for Scientific Research A (18H04094) and C (21K11826).

REFERENCES

[1] Meyer Bertrand: “On formalism in specifications,” IEEE Software, vol. 2, no. 1, pp. 6–26 (1985).

[2] Hiroya Ii, Kozo Okano, and Shinpei Ogata: “Improving Accuracy of Automatic Derivation of State Variables and Transitions from a Japanese Requirements Specification,” Joint Conference on Knowledge-Based Software Engineering Springer, pp. 20-34 (2020).

[3] atilika: “Open source Japanese morphological analysis engine developed in Java,” https://www.atilika.com/ja/kuromoji/ (2022.6.7 access)(in Japanese).

[4] MDN Web Docs, “XPath,” https://developer.mozilla.org/ja/docs/Web/XPath/ (2022.1.20 access).

[5] NuSMV, “An overview of NuSMV,” https://nusmv.fbk.eu/NuSMV/ (2022.6.7 access).

[6] Edmund M. Clarke, Orna Grumberg, Daniel Kroening, Doron Peled, and Helmut Veith: “Model Checking, second edition (Cyber Physical Systems Series),” The MIT press (2018).

[7] Tadao Kasami: “An efficient recognition and syntax-analysis algorithm for context-free languages,” Coordinated Science Laboratory, pp. 1-46 (1965).

[8] SESSAME, “Instructional Materials for Embedded Systems GOMA-1015 Type Requirement Specification,” https://www.sessame.jp/workinggroup/WorkingGroup2/ (2021.3.29 access) (in Japanese).

[9] Masanosuke Ohto, Hiroya Ii, Kozo Okano, and Shinpei Ogata: “Proposal of Extracting State Variables and Values from Requirement Specifications in Japanese by using Dependency Analysis,” 25th International Conference on Knowledge Based and Intelligent information and Engineering Systems, pp. 1649-1657 (2021).

[10] CaboCha: “Yet Another Japanese Dependency Structure Analyzer,” https://taku910.github.io/cabocha/ (2022.10.8 access)(in Japanese).

[11] Taku Kudo, Yuji Matsumoto: “Japanese Parsing by Stepwise Application of Chunking,” IPSJ, Vol43, No.6, pp. 1834-1842 (2002)(in Japanese).

[12] Standard Development Organization: “SBVR,” https://www.omg.org/spec/SBVR/1.5/About-SBVR (2022.1.24 access).

[13] Hina Afree and Imran S. Bajwa: “Generating UML class models from SBVR software requirements specifications,” 23rd Benelux Conference on Artificial Intelligence, pp. 23-32 (2011).

[14] Imran S. Bajwa, Mark G. Lee, and Behzad Bordbar: “SBVR business rules generation from natural language specification,” AAAI 2011 Spring Symposium Series (2011).

[15] Masakazu Takahashi, Satoru Takahashi and Yoshikatsu Fujita : “A Proposal of Adequate and Efficient Designing of UML Documents for Beginners,” Knowledge-Based Intelligent Information and Engineering Systems, KES 2007, vol. 4693, pp. pp 1331–1338 (2007).

[16] Maiko Onishi, Hiroya Ii, Shinpei Ogata, and Kozo Okano: “Extraction method for transition relations from conditional statements in natural language requirements specifications,” Technical Report of IEICE, vol. 2021-SE-208, no. 6, pp. 1-6 (2021) (in Japanese).

[17] Hiroya Ii, Masanoseke Ohto, Hitoshi Kiryu, Shinpei Ogata, and Kozo Okano: “Proposal of a form of requirement specifications for automatic transition model derivation and the derivation method,” Technical Report of IEICE, vol. 121, no. 94 , pp. 13-18 (2021) (in Japanese).

[18] Alessandro Cimatti, Edmund Clarke, Fausto Giunchiglia, and Roveri Marco: “NuSMV : a new symbolic model checker,” International Journal on Software Tools for Technology Transfer2, pp. 410-425 (2000).

Session 2:
Systems and Services
(Chair: Takuya Yoshihiro)

Design of Air Flow Channel for the Droplet Spray Type Olfactory Display

Yohei Seta[*,***], Mitsunori Makino[**], Yuichi Bannai[***], and Motofumi Hattori[***]

[*]Graduate School of Science and Engineering, Chuo University, Japan
[**]Faculty of Science and Engineering, Chuo University, Japan
[***]Department of Information Media, Kanagawa Institute of Technology, Japan
{a15.s845, makino.fme7}@g.chuo-u.ac.jp
{bannai, hattori}@ic.kanagawa-it.ac.jp

Abstract -To date, olfactory displays based on various principles have been developed to present controllable scents.

In this study, to design a novel piezoelectric olfactory display, we conducted experiments to determine the appropriate channel height at which the bottom of the air flow channel does not come in contact with the sprayed purified water; in addition, we obtained the appropriate shape of the connection between the liquid fragrance atomizing mechanism and the air flow channel.

Accordingly, it was inferred that when the distance from the piezoelectric element surface to the bottom of the air flow channel was ≥ 70 mm, contact between the atomized purified water and the bottom of the channel could not be confirmed. It was also deduced that the retention of atomized droplets can be reduced by connecting the airflow channel with a rectangular hole that matches the size of the bottom of the liquid fragrance tank and reducing the distance from the upper surface of the airflow channel to the bottom of the liquid fragrance tank to ≤ 1 mm.

Keywords: olfactory display, pulse ejection, droplet atomization

1 INTRODUCTION

Various forms of olfactory display have been developed for system-controllable scent presentation devices. Bannai et al. developed droplet atomization-type olfactory displays that can be controlled in 100-ms increments, to primarily investigate human responses to olfactory stimuli [1].

A unique feature of droplet atomization olfactory displays is the pulsed presentation of the scents. Pulsed presentation of scents is a method of presenting scents by intermittently atomizing liquid fragrances for a short period (0.1–0.3 s), thereby reducing the amount of atomization and the effects of adaptation and lingering scents.

Nakamura et al. developed an olfactory display using a piezoelectric element as a droplet atomization mechanism, to improve maintainability and simplify the structure of droplet atomization-type olfactory displays. [2]. Nakamura et al.'s piezoelectric olfactory display vibrates its piezoelectric element at 100 kHz when a voltage is applied, and the liquid in contact with the surface of the piezoelectric element can be atomized as tiny droplets from the opposite side through the central group of micropores. This piezoelectric element is attached to the bottom of the liquid fragrance tank and can control the atomization of the liquid fragrance in 1-ms increments.

In principle, the atomized fragrance components may be subject to turbulent diffusion in the process of being transported by air or retention in airflow channel depressions. Seta et al. conducted experiments to validate the performance of Nakamura et al.'s piezoelectric olfactory display [3].

Accordingly, it was verified that the atomized liquid fragrance contacted the bottom of the airflow channel before being fully vaporized, and then remained as droplets.

These problems may result in the continuous presentation of scents independent of the intention of the system, which may severely affect the accuracy of the results obtained from olfactory characterization studies via this system.

In this study, to design novel piezoelectric olfactory displays, experiments were conducted to determine the air flow channel height at which the bottom of the air flow channel does not come into contact with the atomized liquid fragrance. The condition of the atomizing droplets was captured using a digital camera from the side, and the presence or absence of droplet adhesion was verified from the video. We checked for atomization by altering the height of the floor and confirmed the minimum distance at which the droplets do not adhere to the floor.

A simulation also measures the amount of the atomized substance exiting from the air flow channel with a delay from the system's intended scent presentation time due to entrapment of the atomized substance in the vortex generated by the depression in the connection between the droplet atomization mechanism and air flow channel.

Accordingly, it was deduced that when the distance from the surface of the piezoelectric element to the bottom of the channel was ≥ 70 mm, contact between the atomized purified water and the bottom of the air flow channel was not observed.

Instead of the conventional method of connecting the liquid fragrance tank and the air flow channel with a circular hole, a rectangular connection was made to match the size of the bottom of the liquid fragrance tank, and the distance between the upper surface of the air flow channel and the bottom of the liquid fragrance tank was reduced to ≤ 1 mm, thereby reducing the retention of the atomized substance due to vortex in the depression that connects the liquid fragrance tank and the air flow channel.

By designing the air flow channel of a piezoelectric olfactory display that considers these two results, the presentation of scents is expected to be possible in a time closer to the intention of the system than with conventional devices.

2 METHODOLOGY

2.1 Olfactory Display

Among the devices that present stimuli to the five human senses, those that present visual stimuli are called displays, while those that present olfactory stimuli are called olfactory displays. Odors are sensed when odorants with a molecular weight of less than approximately 300 are captured from the air by the olfactory receptor cells in the olfactory epithelium of the nasal cavity. Olfactory displays are artificial scent presentation systems that provide a mechanism that delivers fragrance components to the user's nose by volatilizing liquids such as essential oils and synthetic fragrances, which are sources of the scent.

In olfactory displays, the basic functional components include a volatilization method of liquid fragrance, a concentration control method of volatilized fragrance combined with a mixing method of multiple fragrances, and a time control method to determine the presentation time and when the odor is delivered to the user; in addition, various methods have been proposed [4].

2.2 Inkjet Type Olfactory Display

Bannai et al. developed an inkjet olfactory display using an inkjet printer cartridge as the droplet atomization device.

The inkjet olfactory display employs the nozzle of a commercially available inkjet printer to eject tiny droplets of a solution containing a certain concentration of liquid fragrance, instead of ink. A fan installed at the rear of the device blows air to vaporize microdroplets in the air and directs fragrance components to the user's nose, thereby presenting a scent to the user. The microdroplet ejection can be controlled in 0.1-s increments, and 255 ejection heads can simultaneously eject droplets. The number of microdroplet ejections can be modified in the application, and the ejection volume can be altered in 256 steps from 0 to 255. Each ejection head can eject approximately 4.7 pL droplets at intervals of 1–150 ejections per 0.1 s. Hence, the amount and timing of the liquid fragrance ejection can be fine-tuned, thus enabling the pulse presentation of scents by repeatedly presenting minute droplets of liquid fragrance for significantly short durations.

It has been reported that this method reduces the challenges posed by olfactory adaptation, in which the user becomes less sensitive to smells when continuously exposed to odor stimuli, including the problem of lingering smells, in which smells remain in space and mix when the olfactory display is continuously used, compared to olfactory displays that continuously present scents [5]. By presenting scents to subjects with varying intensities and high accuracy, it was possible to determine the detection threshold, which is the minimum intensity at which a person perceives the presence of a certain scent.

Experiments were conducted using this olfactory display to measure human olfactory characteristics, and revealed that it is possible to perceive scents without adaptation for more than one minute by presenting pulses of scents, that the detection threshold does not change during inhalation at the beginning, in the middle, or at the end of inhalation, and that the human olfactory system can perceive scents from the beginning to two thirds of the inhalation time within two seconds, the average inhalation time in a person at rest [5].

Furthermore, a study was conducted to investigate the interaction between olfaction and vection, a self-moving sensation derived from vision, using an inkjet olfactory display. The obtained results strongly suggested that vection stimulation influences olfaction [6].

However, the problem with this olfactory display is that its structure is complicated and it is not easy to clean. It has been verified that almost all the atomized fragrance droplets vaporize, but the fragrance adheres to the connection between the ink cartridge and the airflow channel, as well as to the injection port. Furthermore, the principle of droplet ejection is based on the thermal method, in which bubbles are generated in a liquid by heating, and then ejected by the pressure of the bubbles, which limits the use of heat-sensitive fragrances.

2.3 Piezoelectric Olfactory Display

Nakamura et al. developed a piezoelectric olfactory display that adopts a piezoelectric element as a droplet atomization mechanism to circumvent the disadvantages of inkjet olfactory displays, such as thermal injection, difficulty in downsizing the atomization mechanism, and complicated cleaning and maintenance [2].

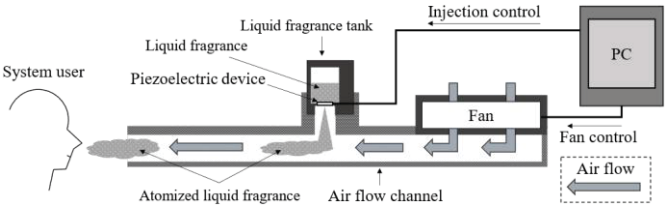


Figure 1: Cross-sectional image of piezoelectric olfactory display (side).

Figure 1 illustrates the operation of the piezoelectric olfactory display. The device uses a fan to blow the liquid fragrance atomized by a piezoelectric device built into a liquid fragrance tank, to present a scent to the user. The piezoelectric device has a diameter of approximately 1 cm and 100 micropores of 9-μm diameters drilled in its center. Although the liquid fragrance in the liquid fragrance tank does not pass through the micropores owing to surface tension when no voltage is applied, it is atomized through the micropores by vibration when voltage is applied. The piezoelectric device oscillated once every 10 μs, and each oscillation could control the atomization at approximately 167 pL. A fan was employed to blow air over the atomized liquid fragrance, promote the volatilization of the liquid fragrance, and present the scent to the user.

A 3-mL capacity liquid fragrance tank was placed above the piezoelectric device, and the fragrance in the tank was continuously ejected from the piezoelectric device. The tank, piezoelectric device, and piezoelectric device holder at the bottom of the tank were removable for easy cleaning and maintenance.

An experiment to measure the mass of purified water atomized 10 times for 10 s verified that the average injection mass was 0.17 g (standard deviation: 0.0077) and remained constant regardless of the number of atomization. Different types of fragrances have different viscosities, surface tensions, and other characteristics; hence, the fragrance solutions to be atomized in olfactory displays must be adjusted for their composition.

2.4 Limitations of the piezoelectric device olfactory display

Nakamura et al.'s olfactory display air flow channel has a cross section 20-mm high and 60-mm wide, which follows the structure of the inkjet olfactory display. To install the liquid fragrance tank, a 10-mm rise section was provided at the top of the air flow channel, and a cylindrical hole with a diameter and height of 15 mm and 15 mm, respectively, was made in the rise. The surface of the liquid fragrance tank was attached to the top of the raised section, and atomization was directed vertically toward the bottom of the air flow channel. To prevent the piezoelectric device from being directly touched by hands, a 5-mm high dimple exists from the bottom of the liquid fragrance tank, as a frame to protect the surface of the piezoelectric device, and a space for fixing the piezoelectric device exists above the frame. Therefore, the distance from the atomization position of the piezoelectric device to the bottom of the channel is 40 mm.

This is equivalent to the atomized height of an inkjet olfactory display, and previous experiments using inkjet olfactory displays reported that the atomized substance was sufficiently vaporized in the air flow channel [1]. Therefore, it was assumed that the atomized substance, used in the experiment, in the piezoelectric olfactory display vaporizes immediately after ejection and does not remain as droplets on the wall surface.

However, the atomization mechanisms of the inkjet and piezoelectric device olfactory displays are different, including the composition of the liquid fragrance solution; hence, the vaporization of the atomized substance is also expected to be different.

Seta et al. created an acrylic plate air flow channel with a height of 20 mm and a width of 60 mm to confirm the conditions in the air flow channel immediately after the atomization of droplets in the piezoelectric device olfactory display of Nakamura et al. They connected a liquid fragrance tank of Nakamura et al.'s method to the acrylic plate air flow channel and observed the inside of the air flow channel by capturing video images of the inside of the air flow channel at the time of atomization. The results showed that the atomized substance adhered to the bottom surface of the conventional structure as droplets at the time of atomization of a banana liquid fragrance and purified water and evaporated later than after the system had finished atomizing them.

A cylindrical cavity exists at the connection between the liquid fragrance tank and the air flow channel to maintain more distance between the bottom of the air flow channel and piezoelectric device.

However, it is known that the cavity shape facing the flow triggers unstable flow owing to the contact between fluids with significantly different velocities, thereby resulting in vortex-like flow [7].

When such vortexes are generated, fragrance components may be retained because of the vortexes. This may cause the fragrance components that remain in the cavities to gradually leak out and reach the system user, even after the end of atomization. This problem has not been considered in conventional piezoelectric olfactory displays.

To accurately determine a person's scent detection thresholds using a droplet atomizing olfactory display, it is necessary to switch between odorless and attached states with a high degree of accuracy [8]. If the presentation state by the system is odorless but the subject is actually presented with a scent because of the above problem, the olfactory characteristics of the subject may not be measured correctly.

3 METHOD

In this study, we experimentally determined the height of air flow channel where no droplet adhesion is observed on the bottom of the air flow channel when liquid fragrance is injected in a piezoelectric olfactory display.

Simulations were also performed to confirm vortex generation and retention of the atomized substance in the depression at the connection between the fragrance atomizing mechanism and air flow channel and to determine the shape of the connection to reduce vortex generation and retention.

3.1 Experiment: Confirmation of the required air flow channel bottom distance during purified water atomization

To determine the minimum distance at which a droplet is atomized by the piezoelectric device without contacting the bottom of the airflow channel, an experimental airflow channel was constructed using acrylic plates. Figure 2 presents the configuration and dimensions of the experimental airflow channel.

In the experiment, the distance from the piezoelectric device to the bottom of the airflow channel was varied, and a video of the droplet atomization was captured. The video images of the atomized droplets were observed and checked for the presence or absence of droplet adhesion to the bottom of the air flow channel. The threshold value of the bottom height for the presence/absence of deposits at the bottom of the airflow channel is defined as the minimum distance without contact with the airflow channel bottom when atomizing droplets.

Previous experiments have demonstrated that liquid fragrances containing ethanol or other solvents vaporize faster than purified water, and by determining the threshold of the bottom height relative to purified water, it is possible to consider liquid fragrances that may be used in the future [3]. Therefore, only purified water was utilized in this experiment.

The experimental airflow channel employs a bottom surface whose height can be modified as desired, which is referred to as a movable floor in this study. The movable floor was an

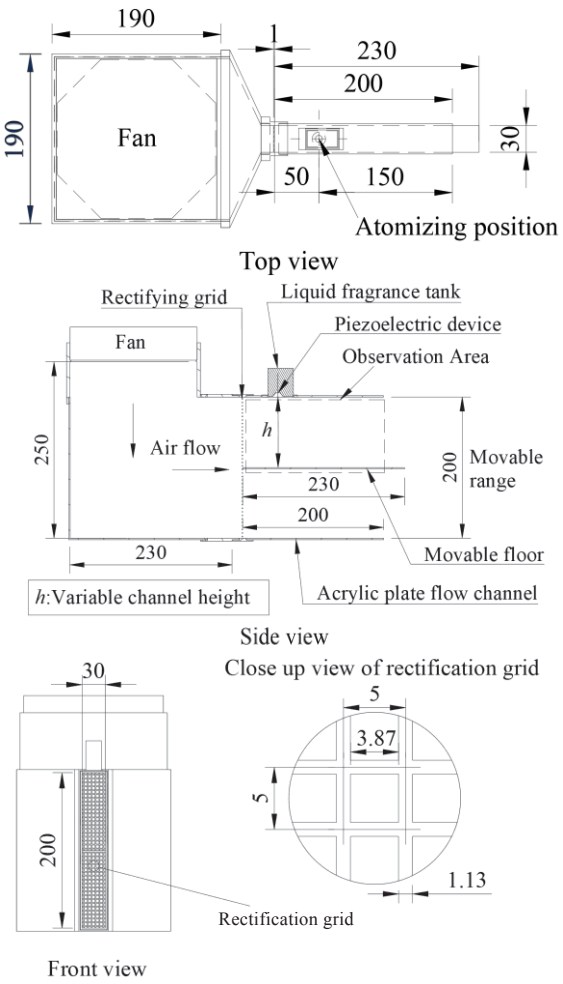


Figure 2: Structure of experimental channel (unit: mm).

acrylic plate with the same dimensions as the bottom of the airflow channel, wrapped with vinyl tape at both ends, and fixed in position by friction with the acrylic plate on the channel side. Therefore, the movable floor can be easily removed when changing the bottom height.

To maintain constant conditions on the movable floor surface during continuous injection, each time the bottom height was changed, the surface was wiped with a paper rag moistened with purified water, and then the surface was wiped dry again with a dry paper rag. When fixing the movable floor, the height of the installed surface was checked with a digital caliper, and a level was used to check the level at the part protruding from the channel before conducting the experiment.

The air velocity in the air flow channel during the droplet atomization experiment was 1.8 m/s, which has been adopted in conventional experiments on inkjet olfactory displays [1]. A PWM-controllable PC fan was utilized to provide a constant airflow in the experimental airflow channel. The selection of the fan to be used is based on the maximum airflow rate, and the fan is required to sufficiently satisfy the target air velocity over the entire experimental air flow channel cross section. Blowing air by rotating fans generates a low velocity near the center of the air velocity distribution, asymmetrical bias, and turbulence. To improve the reproducibility of the experimental results, it was necessary to stabilize the airflow in the experimental airflow channel.

In this experimental air flow channel, a grid-like structure used for rectification in the experimental wind tunnel was installed in the cross section of the air flow channel that flows into the air flow channel from the fan mounting area. Previous experiments have demonstrated that the grid shape is effective in stabilizing the airflow in the air flow channels of piezoelectric olfactory displays when the pitch width, plate thickness, and open area ration are 5 mm, 1 mm, and 60%, respectively [9]; in addition, the same parameters will be adopted in this study. The fan was mounted in the direction of intake from the top of the experimental apparatus and perpendicular to the cross-section of the air flow channel, to reduce the influence of the low-speed portion at the center of the fan.

During the experiment, an anemometer was placed at the center of the cross section, which is at the end of the experimental air flow channel, and the power of the fan was adjusted to achieve the target wind velocity while checking the wind velocity. The experiment was allowed only between wind speeds of 1.8 ±0.03 m/s.

3.1.1. Preliminary experiment: Confirm of wind velocity distribution in the experimental air flow channel

Table 1: Average wind velocity of channel cross section

	Horizontal (h = 100 mm)	Vertical (h = 100 mm)	Horizontal (h = 200 mm)	Vertical (h = 200 mm)
Average wind velocity (m/s)	2.61	2.70	2.80	2.70
Standard deviation	0.21	0.12	0.21	0.06

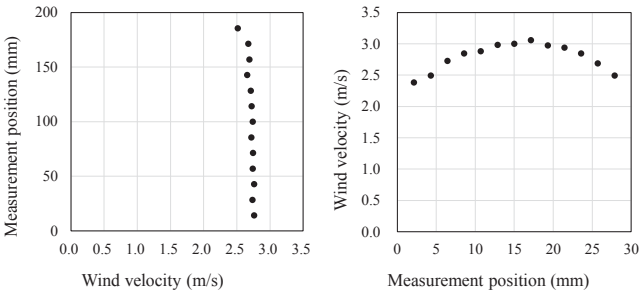


Figure 3: Wind speed distribution (h = 200 mm).

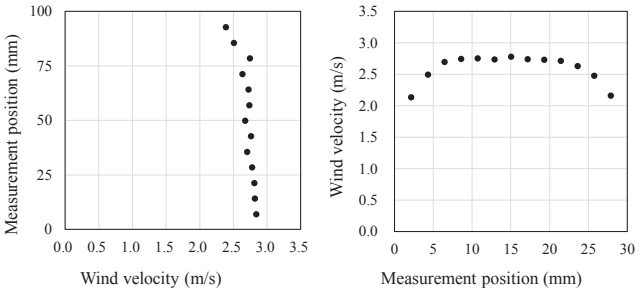


Figure 4: Wind speed distribution (h = 100 mm).

To verify the maximum air velocity and the rectification effect in the experimental air flow channel and to check

whether the use of the movable floor caused a significant disturbance of the air velocity distribution in the air flow channel cross section, the air velocity distribution was measured at the vertical and horizontal centers of the cross section at the airflow channel end when the experimental air flow channel height was h = 200 mm and h = 100 mm with the movable floor at the output under the maximum airflow rate of the fan.

In both cases, the vertical wind velocity distribution was almost flat, with an average of approximately 2.7 m/s, and the horizontal wind velocity distribution was almost symmetrical, with the fastest near the center and gradually slowing down as it approached the wall (Figures 3 and 4). Table 1 presents the average wind velocity and standard deviation of the cross section of the experimental air flow channel.

The obtained results indicate that the experimental air flow channel can be utilized for this experiment because the target air velocity of 1.8 m/s is sufficiently satisfied, and negligible turbulence exists in the air velocity distribution owing to the asymmetric flow caused by the fan rotation and the influence of the movable floor.

3.2 Simulation: Confirmation of vortex generation and retention of atomized substance owing to the shape of the fragrance tank connection

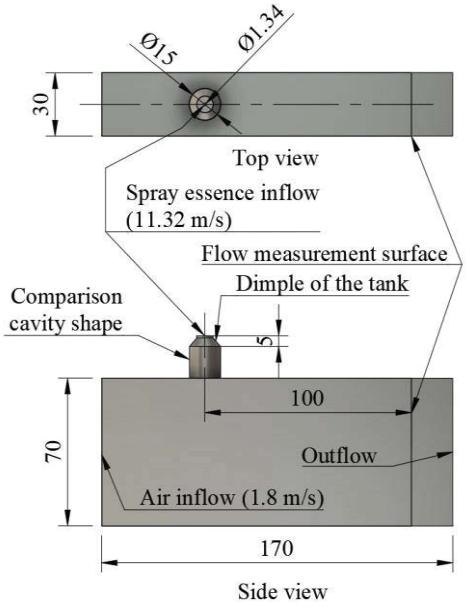


Figure 5: Simulation environment (unit: mm).

To obtain an appropriate shape for the connection between the liquid fragrance tank and the airflow channel, we checked the occurrence of vortexes inside it and compared the effect of its shape on the retention of the atomized substance. The shape of the connection between the air flow channel and liquid fragrance tank cannot be directly observed because the material of the liquid fragrance tank does not allow light to penetrate under the current experimental environment. If the atomized substance remains inside, the droplets will evaporate inside it and will be invisible. Therefore, in this study, the motion of the atomized substance was numerically calculated via fluid simulation, using the particle method to predict the internal state of the atomized substance.

The particle method is a computational technique that is mainly adopted for fluid simulation, in which a continuum is discretely represented as a set of movable particles. The governing equations determine the forces acting on each particle, and the fluid phenomena can be calculated by continuously updating the coordinates of the particles over time. Compared with the lattice method, which describes fluid motion by dividing the space with a fixed grid and calculating the fluid inflow and outflow at the grid points, the particle method has the advantages of not requiring the space to be cut with a mesh and being able to easily track individual fluid elements as they flow in. In addition, by explicitly obtaining all terms of the governing equations, parallel computation is facilitated and benefits from accelerated calculation using GPUs. Therefore, the MPS explicit method, a particle method, is employed in this study to calculate the air flow in an air flow channel [10].

The air in the air flow channel and the atomized substances treated in this study were sufficiently slower than the speed of sound to be treated as incompressible fluids. The Navier-Stokes equation in Equation (1) and the continuity equation in Equation (2) are adopted as the governing equations for incompressible fluids.

$$\frac{D\boldsymbol{u}}{Dt} = -\frac{1}{\rho}\nabla \boldsymbol{P} + v\nabla^2 \boldsymbol{u} + \boldsymbol{F} \qquad (1)$$

$$\frac{D\rho}{Dt} = 0 \qquad (2)$$

where $\boldsymbol{u}$, ρ, $\boldsymbol{P}$, v, $\boldsymbol{F}$, and t denote the flow velocity, mass density, pressure, kinematic viscosity, external force, and time, respectively. Equation (2) refers to the law of conservation of mass in dynamics, and it was adopted to calculate the particle dynamics.

In this study, calculations were performed using air and a vaporized banana liquid fragrance as particles in the MPS method with two parameters. The air particle parameters refer to the JIS standard pneumatic fluid power - standard reference atmosphere, with a humidity of 65%, a temperature of 20 °C, and a pressure of 1 atm as the setting environment [11]. From the setting environment, a density of 1.198 kg/m^3 and kinematic viscosity of 1.505E-5 m^2/s were adopted as parameters for air particles [12].

The parameters of the particles of the atomized substance are strictly a phenomenon of gradual vaporization in air from the liquid state immediately after atomization. However, for computational simplicity, the present simulation was based on the assumption that the particles enter the air in a gas state immediately after atomization. The substance to be atomized was a mixture of 5 % isoamyl acetate, 4.25 % ethanol, 0.94 % glycerin, 0.09 % propylene glycol, and 89.72 % water as banana fragrance, based on the composition ratio of a banana liquid fragrance solution that can be adopted in the piezoelectric olfactory display by Nakamura and colleagues [2]. Based on the molecular weight of each component and the mixing ratio, we obtained an average molecular weight of 25.57 for the banana liquid fragrance mixture. Assuming that all the liquid mixtures with the average molecular weight

were vaporized, the density of the particles to be sprayed was 1.0631 kg/m^3 according to Avogadro's law, which states that 1 mol of gas has a volume of 24.04 L at a temperature of 20 °C and 1 atm. The kinematic viscosity was assumed to be 1.505E-5 m^2/s, the same as that of air particles. Figure 5 illustrates the computational environment.

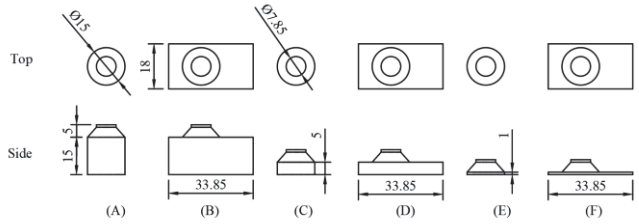


Figure 6: Comparison cavity shape (unit: mm).

The computational air flow channel has an inflow of air particles in an air flow channel cross section with a width of 30 mm and a height of 70 mm, and the initial velocity was set to 1.8 m/s. The computational air flow channel length was 170 mm, the center of the inflow point of the particles of the atomized banana liquid fragrance was 50 mm from the inflow point of air, and the cross section of the computational air flow channel at 100 mm from the inflow point was the surface from which the flow rate of the particles of atomized banana liquid fragrance was measured.

Figure 6 presents the shape of the cavity at the connection between the liquid fragrance tank and the airflow channel in the comparison model. The area of the trapezoidal rotator at the top of the cavity shape in Figure 6 represents the dimple in the frame protecting the piezoelectric device of the liquid fragrance tank, which was 5 mm high and fixed in all cases, with its top surface in contact with the atomizing surface of the piezoelectric device. The variables include the height of the cylindrical areas that are in contact with the dimple of the liquid fragrance tank and the height of the rectangular area that matches the size of the bottom of the liquid fragrance tank, Thereby creating a cavity model with heights of 15 mm, 5 mm, and 1 mm for the cylindrical and rectangular areas, respectively. The cavity shapes of the six patterns were compared.

The initial inflow velocity of banana liquid fragrance particles is calculated from the distance traveled by the piezoelectric device during the atomization of purified water, which was recorded as a 240-fps movie, at the frame when the atomization is first observed, and in this calculation, it was 11.322 m/s. Banana liquid fragrance particles flowed in a vertical direction from the center of the depression in the fragrance tank toward the bottom of the flow path. The banana liquid fragrance particles inflow was 0.3 s from 0.5 s, to 0.8 s from the start of the simulation, which represents a single pulse of injection. The calculation covered a period of up to 2.0 s, which is more than 1.0 s after the end of the inflow of banana liquid fragrance particles.

The interior of the cavity shape was observed as the banana liquid fragrance particles flowed and vortex generation was checked. Banana liquid fragrance particles passing through the flow measurement surface were counted to obtain the time series variation of the volume flow rate per second. Furthermore, the total volume of the target particles passed during the 1-s period from 0.883 s to 1.883 s after the end of spraying, when the banana liquid fragrance particles were fully discharged from the channel without retention, was compared with the volume of the target particles that flowed late due to retention. Based on the above results, the most suitable shape for olfactory display among the shapes to be compared was considered.

4 RESULT

4.1 Experiment: Confirmation of the required air flow channel bottom distance at the time of purified water atomization

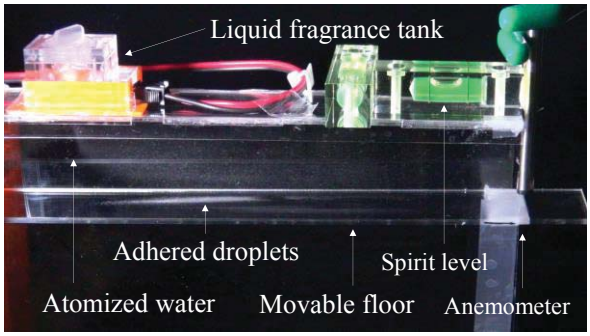


Figure 7: Illustration of experimental results (h = 34 mm).

An experimental air flow channel was fabricated by 3D printer and acrylic sheet fabrication. Figure 7 illustrates the atomization at h = 34 mm. A digital camera was positioned at the center of the screen at an angle that captured the entire movable floor surface, and any droplet adhesion on the bottom of the channel during purified water spraying was illuminated by the white LED light adopted as illumination and visualized in white.

Figures 8, 9, and 10 present the experimental results for h = 34, 63, and 64 mm, respectively, during 1 s of purified water atomization, and the time series changed with the atomization start frame set to 0 s. These images were clipped only inside

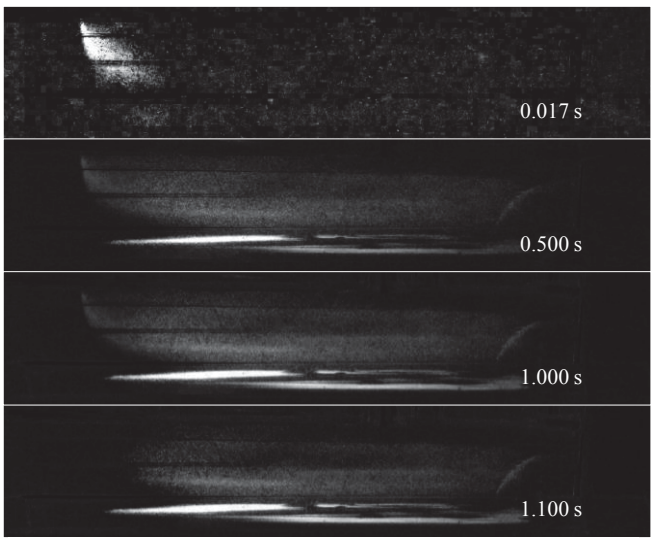


Figure 8: Experimental results (h = 34 mm, Color difference from 0 s and after contrast adjustment, Clipping only in air flow channel).

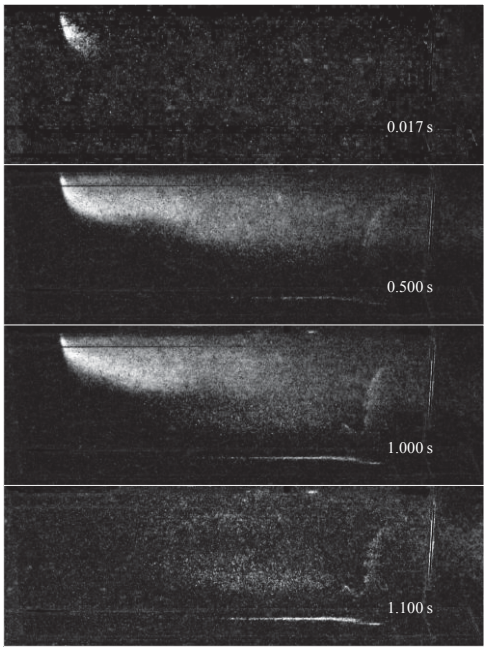


Figure 9: Experimental results (h = 63 mm, Color difference from 0 s and after contrast adjustment, Clipping only in air flow channel).

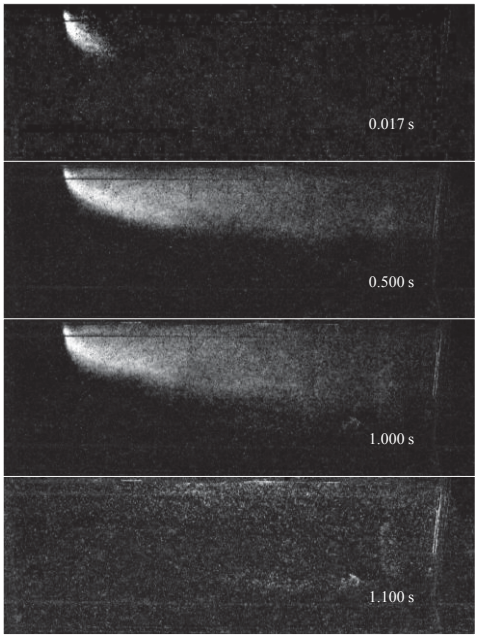


Figure 10: Experimental results (h = 64 mm, Color difference from 0 s and after contrast adjustment, Clipping only in air flow channel).

the airflow channel, and the difference in color values between the atomizing start frame and each frame was created using Adobe Photoshop CS5. The color values were output only for the areas that changed from the initial frame, and the areas that did not change are output in black.

At h = 34 mm in Figure 8, droplet adhesion is clearly observed over a wide area of the movable floor surface. Droplets remained at the bottom of the air flow channel for 1.1 s after the atomization was completed and the atomized substance had flowed out of the air flow channel. In Figure 9, at h = 63 mm, linear droplet adhesion is observed in the center of the movable floor in front of the air flow channel in the form of a thin line. In Figure 10, h = 64 mm, the atomized substance flowing in the channel was confirmed, as illustrated in Figure 9; however, no bottom adhesion was observed, and the same result was obtained in a total of six subsequent trials conducted for confirmation.

4.2 Simulation: Confirmation of vortex generation and retention of atomized substance due to the shape of the fragrance tank connection

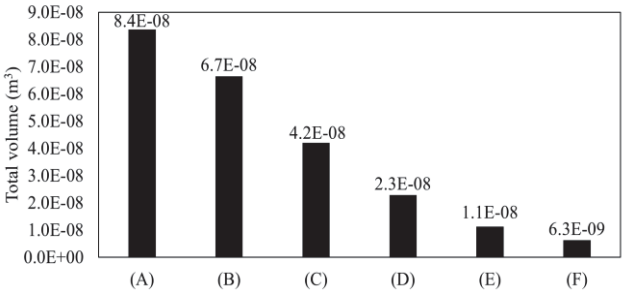


Figure 11: Total volume passing through the measurement surface up to 1 s after the end of atomization.

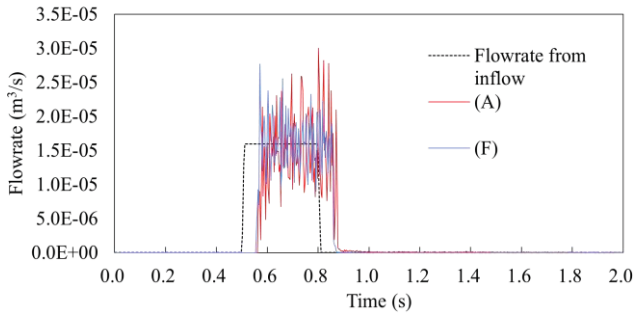


Figure 12: Simulation results Comparison of flow rate through shape (A) and (F).

Based on the computational environment set up, Prometech Particleworks 7.1 was employed as the simulation software for the calculations. The computing environment was a CPU: Intel core i9-10900K, memory:128 GB, OS: Microsoft Windows 10 Pro, and GPU: Nvidia RTX 8000. The computation time was approximately 30–35 h for each case in 2 s of real time. The number of particles in the entire channel was approximately 10 million.

The calculation results were observed; in all cases, a vortex-like flow was observed in the cavities connecting the liquid fragrance tank. Figure 11 compares the total volume of the atomized substance passing through the measurement area between 0.883 and 1.883 s after the end of the atomization. The result of (A), a circular shape with a height of 15 mm, which is equivalent to that of the conventional device, exhibited the highest delayed flow, while the result of (F), a rectangular shape with a height of 1 mm, exhibited the lowest delayed flow. At the same height, the rectangular shape tended to have slightly less delayed flow than the circular shape. Figure 12 presents the time series variation of the volume of the atomized substance passing through the measurement area of geometry (A), (F), where the dotted line represents the inflow volume of the atomized substance.

5 ANALYSIS

5.1 Experiment: Confirmation of the required air flow channel bottom distance at the time of purified water atomization

The air flow channel height of the conventional piezoelectric olfactory display (h = 34 mm) was deduced to be insufficient because of the large amount of droplet adhesion on the bottom surface. Bottom adhesions were observed up to h = 63 mm, but no bottom adhesions were observed above h = 64 mm, even after repeated atomization. The required air flow channel bottom distance at the time of purified water atomization under a flow velocity of 1.8 m/s, when vibrating at 100 kHz using a piezoelectric device with 100 micro-holes of 9 μm, was verified to be 70 mm, including the dimple in the liquid fragrance tank.

In all cases, contact was observed on the sides during atomization; therefore, the 30-mm air flow channel width should be reviewed in the future.

5.2 Simulation: Confirmation of vortex generation and retention of atomized substance due to the shape of the fragrance tank connection

The simulation results demonstrated that the delayed flow rate of the atomized substance from the 20-mm high circular cavity was the largest, whereas the delayed flow rate from the 6-mm high rectangular cavity was the smallest. It was also inferred that, for cavities of the same height, the delay flow rate tended to be smaller for the rectangular cavity than for the circular cavity. Extending the cavity in the connection to maintain the distance between the piezoelectric device and the bottom of the airflow channel is not considered good practice in the design of piezoelectric device olfactory displays.

This suggests that, among the shapes compared in this study, the 6-mm high rectangular cavity is the most suitable for the shape of the connection of the piezoelectric olfactory display. The shape of the fragrance tank adopted in this study had a 5-mm high dimple to protect the surface of the piezoelectric device; however, this shape was also inferred to be a factor in the generation of internal vortexes. From this perspective, there is room for further improvement in the shape of the outer frame of a liquid fragrance tank.

6 CONCLUSION

Purified water atomization experiments were conducted to design a piezoelectric olfactory display by varying the bottom distance. To determine the appropriate cavity shape for connecting the liquid fragrance tank to the airflow channel, fluid simulations were conducted to determine the presence of vortex flow and its effect on the flow.

The obtained results demonstrated that when purified water was atomized perpendicularly to the bottom surface for 1.0 s from a piezoelectric device with 100 micropores of 9 μm in the center, and when air flow was present in the channel at an air speed of 1.8 m/s toward the outlet, the distance from the atomized position to the bottom surface was ≥ 70 mm, and the contact of the atomized droplets to the bottom surface was not confirmed.

The obtained simulation results indicate that if a cavity exists at the connection between the droplet atomization mechanism and the air flow channel, a vortex may be created inside, which may cause the atomized droplet to remain in the air flow channel even 1 s after the end of injection. This suggests that, among the shapes compared in this study, the 6-mm high rectangular cavity is the most suitable for the shape of the connection to the piezoelectric olfactory display. In the future, we will create a novel air flow channel structure designed based on the results of this study. Furthermore, we will conduct experiments on human subjects to verify the effectiveness of the piezoelectric device olfactory display in experiments and investigate the characteristics of human olfaction.

The piezoelectric olfactory display with a novel air flow channel structure designed considering these two results is expected to reduce the difference between the time when fragrance components actually reach the subject and that set by the system, compared to the conventional model.

The effectiveness of this method will be verified through experiments on test subjects and measurements using gas sensors.

REFERENCES

[1] A. Kadowaki, J. Sato, Y. Bannai, and K. Okada, Measurement and Modeling of Olfactory Responses to Pulse Ejection of Odors, J. Japan Association on Odor Environment, Vol.39, No.1, pp.36-43 (2008).

[2] S. Nakamura, and Y. Bannai, Development of An Olfactory Display Using a Piezoelectric Element and Measurement of Olfactory Detection Threshold, VRSJ Research Report. Vol.25, No.SBR-1, pp.1-6 (2020).

[3] Y. Seta, M Makino, Y. Bannai, and M. Hattori, Proposal of Air-Flow Channel Structures of The Olfactory Display Using the Piezoelectric Element Considering Fluid Behavior, VRSJ Research Report, Vol.27, No.SBR-1, pp.9-14 (2022).

[4] Y. Bannai, "Olfactory display using piezoelectric element," Odor sensing, analysis and its visualization and quantification, pp.518-529, Technical Information Institute, ISBN 978-4-86104-810-4 (2020).

[5] J. Sato, A. Kadowaki, K. Ohtsu, Y. Bannai, and K. Okada, Scent Presentation Technique by Pulse Ejection to avoid Olfactory Adaptation, Transactions of Information Processing Society of Japan, Vol.49, No.8, pp.2922-2929 (2008).

[6] A. Aruga, Y. Bannai, and T. Seno, Investigation of the Influence of Scent on Self-Motion Feeling by Vection, International Journal of Informatics Society, Vol.11, No.2. pp.65-73 (2019).

[7] K. Fujita, F. Takahashi, T. Yonemura, N. Kawabata, and T. Ohta, Study of the Flow Structure of a Newtonian Fluid through a Rectangular Channel with a Cavity,

Research reports of National Institute of Technology, 50th Anniversary Issue, No.49, pp.1-8 (2015).
[8] A. Aruga, S. Nakamura, and Y. Bannai, Fine Droplet Discharge Type Olfactory Display, 24th Annual Conference of the Virtual Reality Society of Japan, O-02 (2019).
[9] Y. Seta, N. Mori, M. Makino, Y. Bannai, and M. Hattori, Comparison of Rectification Effects of Grid Shapes in an Air-Flow Channel for Droplet Spray Olfactory Displays, Proceedings of the 84th National Convention of IPSJ, Information Processing Society of Japan, 2F-05 (2022).
[10] M. Oochi, S. Koshizuka and M. Sakai, Explicit MPS Algorithm for Free Surface Flow Analysis, Transactions of the Japan Society for Computational Engineering and Science, Vol.2010, p.20100013 (2010).
[11] Japanese Industrial Standards: Pneumatic fluid power− Standard reference atmosphere, JIS B 8393: 2000 (2000).
[12] National Astronomical Observatory of Japan, Chronological Scientific Tables 2019, Maruzen (2019).

Proposal of Potted Flower Area Segmentation Method Using Hough Transform

Shun Fujii†, Yoshitaka Nakamura‡, Hiroshi Inamura*, and Shigemi Ishida*

†Graduate School of Systems Information Science, Future University Hakodate, Japan
‡Faculty of Engineering, Kyoto Tachibana University, Japan
* School of Systems Information Science, Future University Hakodate, Japan
{g2121048, inamura, ish}@fun.ac.jp, nakamura-yos@tachibana-u.ac.jp

Abstract -
The impact of plant diseases on agricultural production is significant. Therefore, plant diseases must be removed or otherwise treated as soon as they are detected. Conventional methods have attempted to improve the accuracy of disease detection. However, the current situation is costly in terms of collecting a large amount of image data for disease detection, installing and managing monitoring equipment, and extracting disease images. In this study, we propose a method to segment an image of the entire farm field using image processing based on the Hough transform to isolate individual plants. We aim to improve the accuracy of disease detection by performing disease identification on cropped plant images using the proposed method. In the evaluation, we confirmed that the proposed method can trim a single plant with an F-measure of 72.5%, which is approximately 7 percentage points higher than that obtained by segmenting plant parts using CNNs.

Keywords: smart agriculture, plant disease, image processing, Hough transform

1 Introduction

Zadoks et al. reported that more than one-third of agricultural production is wasted due to plant diseases [1]. Since plant diseases are spread by infection among plants, it is necessary to remove or otherwise treat plant diseases as soon as they are detected.

Several types of research conducted on this issue, aiming at automating the prediction and detection of plant diseases. Several methods using deep learning based on Convolutional Neural Networks (CNNs) have been proposed. These researches include the detection of cucumber and tomato diseases [2], [3]. In addition to vegetables, there are also researches to detect diseases of rice plant [4].

Since these researches target the plant disease itself, they require data on specific areas such as diseased leaves. Therefore, more than 10,000 images need to be collected, and the cost of installing and maintaining monitoring equipment and extracting images of diseased areas cannot be ignored. Suwa, et al. propose a method for detecting diseases by extracting plant leaves from images of the entire field [3].

As described above, although disease detection has been conducted for specific parts such as diseased leaves, but the research on disease detection from images of the entire farm has not been established. Therefore, we propose to construct a system that can extract specific parts of a field, such as leaves, from the entire farm field image.

In our study, we focus on potted flowers, which are arranged in a grid pattern and thus each plant is easily extracted. We propose a 2-step disease detection method, in which each plant is extracted from the entire farm field image by segmenting the potted flower area, and then each plant is identified for disease identification. In this paper, we consider a method for potted flower segmentation that uses feature extraction of grid patterns that appear in the field image.

This paper is organized as follows. Section 2 reviews related studies on plant disease detection, and Section 3 describes the proposed method for detecting plant diseases from images of the entire farm field. Section 4 evaluates the proposed method and discusses the evaluation results, and Section 5 concludes the paper.

2 Related Studies

Methods for detecting diseases in vegetables can be classified into 2 categories: methods using deep learning based on CNNs, and methods using image processing.

2.1 Plant Disease Detection Method Using Deep Learning Based on CNNs

CNNs are widely used in image analysis and are also widely used for plant disease detection. Fujita, et al. achieve 83.2% of identification accuracy in the classification of 9 classes of plant diseases on cucumber, including 8 types of diseased leaves and healthy leaves [2]. Since these methods use image data trimmed to a single leaf as training data, it is necessary to collect data on leaves suspected to be diseased. Regarding the leaf trimming problem, Suwa, et al. detected plant diseases from images of the entire farm field and achieved disease detection rate of 91.1% [3]. In this method, a 2-step detection process is used: leaf extraction from the field image, and disease identification on the extracted leaves.

Although the evaluation results of disease detection using CNNs were reported to be generally highly accurate, these methods require a large amount of data, and the cost of data collection is high. About 8,000 images were used in Ref.[2], and about

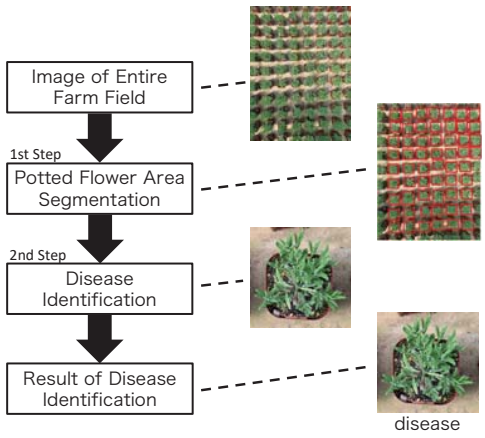


Figure 1: Overview of Proposed System

50,000 images in Ref.[3]. The disease detection rate decreased to 38.9% when images taken in a different farm field from the one used for the training data were applied to the detector [3].

2.2 Plant Disease Detection Method Using Image Processing

Several types of research use image processing to detect diseases with a small number of data. Jadhav, et al. proposed a method to extract only the image area of the disease on leaves [5]. This research also used images of only a single leaf as data like these researches [2], which does not reduce the cost. Kawasaki proposed a method for trimming a single leaf from an image of the entire farm field using image processing [6]. However, this method failed in trimming a leaf image. The reason for this failure is stated to be that the image channels used for feature acquisition are not suitable for leaf detection.

As shown in Ref.[5], it is possible to identify the type and infection level of a disease by extracting only the diseased area if the disease to be detected has prominent features. However, only image processing is difficult to extract leaves from the entire field image. While image processing methods can detect diseases without requiring a large amount of data, features common to all images are required for disease detection.

3 Disease Detection System Using Potted Flower Area Segmentation

In this study, we propose a 2-step system that identifies diseases after individual plants are extracted, to enable disease detection from images of the entire farm field. The proposed system is applied to potted plants, which are arranged in a grid pattern and thus each plant can be easily extracted. Figure 1 shows an overview of the proposed system.

In order to segment potted flower areas using image processing, this method takes advantage of the feature that potted flowers are arranged and grown in a grid pattern over the entire field. Due to this feature, grid lines appear when background images other than potted flowers are extracted. The main idea of this method is to extract each potted flower by trimming using the intersection points of these grid lines. The grid lines that appear between each potted flower are detected using the Hough transform[7]. The process of segmenting the potted flower area using the Hough transform and trimming individual potted flowers is shown in Fig.2.

3.1 Area extraction for applying Hough transform

The pre-processing for detecting grid lines from the farm field images in the proposed system is as follows. Since grid lines appear in background images other than individual potted flower images, this process identifies background images. The proposed method focuses on the green color of the leaf of the individual potted flowers and considers all other colors as the background.

A method for extracting green areas of individual plants use the RGB-based Vegetative Indexes (RGBVI) shown in Expression (1) This Expression (1) makes it possible to extract the features of green areas with high accuracy.

$$\frac{G^2 - (R \times B)}{G^2 + (R \times B)} \tag{1}$$

Our proposed system uses this method to extract green areas from the RGB channels of the entire field image. The image of the entire field has 3 channels, R (red), G (green), and B (blue), and all pixel values are in the range of 0 to 255. After applying Equation(1), the pixel value of -1 to 1 is set to 0 to 255 and converted to a gray-scale image.

The converted gray-scale image is a multi-stage image, and the Hough transform cannot be applied. Therefore, image binarization is required. Our proposed system applies Otsu's binarization algorithm[8]. After binarization, the farm field image is represented as white for individual potted flowers and black for the background. There are white spots in the individual potted flower area and black spots in the background area, which are noise. The proposed method removes these noises by applying expansion processes to the white areas in the image. Since the Hough transform detects white areas as straight lines, the binarized image is inverted to black and white.

This process makes individual potted flower areas and background areas separate from the entire farm field image.

3.2 Line detection by Hough transform

In the image shown in Fig.2(1), the background area is displayed as a linear white area with many line segments and lines mixed in. If the Hough transform is directly applied to the image, many line segments and lines unrelated to potted flowers are also extracted. Therefore, it is difficult to extract the optimal lines for dividing the individual potted flower area and the background

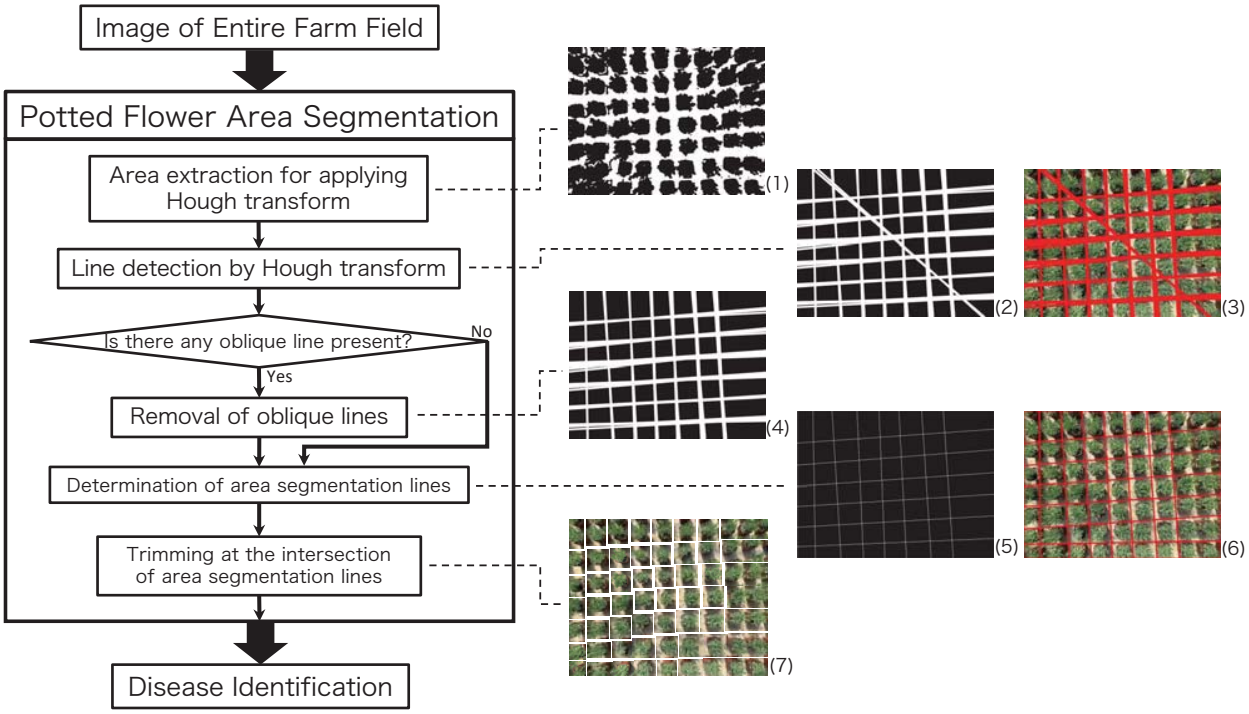


Figure 2: Process of Potted Flower Area Segmentation

area. The proposed system first adjusts the parameters of the Hough transform so that line segments with start and end points in the background area are output. The Hough transform after parameter adjustment is called the segment Hough transform. The segment Hough transform produces line segments that are candidates for line segments that divide the individual potted flower area and the background area. In order to trim the individual potted flower area, these line segments must be straights line that spans the entire image. The overlapping line segments are good candidates for boundaries between areas. Therefore, the line segments outputted by the segment Hough transform are thinned, and then the thinned line segments are expanded. These processes produce straight lines that divide the individual potted flower area and the background area in the entire farm field image. This is called the line Hough transform. When the line calculated here is merged into the input RGB image, it can be seen that a line can be detected in the background.

3.3 Removal of oblique lines

If the image after applying the line Hough transform (Fig.2(2)) contains oblique lines, these lines are noise unrelated to the individual potted flower areas and must be removed. In the process of the Hough transform, the slope and intercept of the detected line are obtained. Based on the slopes, each line is classified using a clustering method to separate the oblique lines. The proposed method uses a density-based clustering method, DBSCAN [9], as the clustering method. Since oblique lines are fewer than straight lines in the horizontal and vertical directions, lines belonging to the top 2 clusters in terms of the number of lines are used as area segmentation lines, and lines belonging to the remaining clusters are removed. If all lines are classified into 2 clusters, no oblique lines are judged.

As an example, the result of clustering the slope of lines for the image shown in Fig.2(2) is shown in Fig.3. Fig.3 explains that the vertical axis represents the slope of the lines, and the horizontal axis is set to 0 as the specified value since the clustering is based only on the slope of the lines. The result shows that the clusters are classified into clusters with 38 lines, clusters with 5 lines, and clusters with 29 lines. The 5 straight lines belonging to the middle cluster are separated as oblique lines and removed (Fig.2(4)).

3.4 Determination of area segmentation lines

Since the image after the line Hough transform has multiple overlapping lines, these lines cannot be used as area segmentation lines for trimming individual potted flowers. Therefore, all the lines are thickened, and the overlapped lines are clustered. If the lines are too thin, they cannot be classified into the same

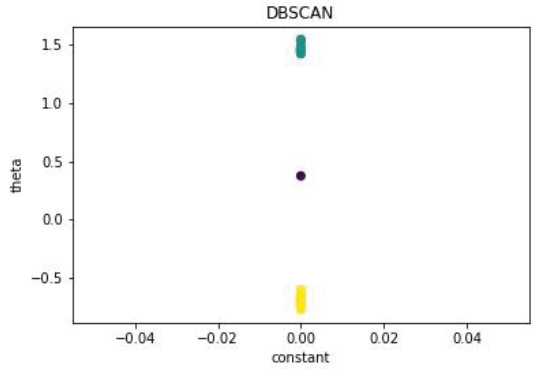


Figure 3: Result of DBSCAN with the Slope of Lines

cluster. If the lines are too thick, lines that should be in different clusters are classified into the same cluster. For this reason, the thickness of the lines should be set to about 20 pixels. The line passing through the center of the clustered lines is used as the area segmentation line (Fig.2(5)). The calculated area segmentation line is found to be the best line for segmenting individual potted flowers(Fig.2(6)).

3.5 Trimming at the intersection of area segmentation lines

Trimming at the intersections of the area segmentation line obtained allows the potted flower to be divided into individual plants. The proposed method trims the largest rectangle with 4 intersections. In this way, the trimmed individual potted flower image retains the maximum amount of image information (Fig.2(7)).

4 Evaluation

We evaluate the potted flower area segmentation method. For the evaluation of the potted flower area segmentation method, we compare the accuracy of the proposed method with that of the CNNs-based segmentation method.

The data used for the evaluation were 78 image. Evaluation data is taken so that "one image shows about 60 plants".

4.1 Evaluation of the Potted Flower Area Segmentation Method

The evaluation criterion of the potted flower area segmentation method is whether or not only one whole plant is trimmed. The term "only one whole plant" here refers to the state in which one whole plant is captured, and only a portion smaller than half of the other plants are captured.

This evaluation criterion are the precision (Equation(2)), the recall (Equation(3)), and the F-measure (Equation(4)).

Table 1: Potted Flower Area Segmentation

	Proposed Method	CNNs-based Method
Precision [%]	78.5	94.3
Recall [%]	67.3	49.5
F-measure[%]	72.5	65.0

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$\text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

In this evaluation, TP (True Positive) is the case where potted flowers are trimmed from the test data image, FN (False Negative) is the case where no potted flower is trimmed from the test data image, and FP (False Positive) is the case where the test data image has been trimmed to include parts of the image that are not potted flowers. Therefore, precision is the probability that the trimmed image is a single potted flower, and recall is the probability that multiple potted flower plants are cropped as a single plant. F-measure is the harmonic mean of precision and recall.

Table 1 shows the results of potted flower area segmentation. The proposed potted flower area segmentation method achieved 78.5% precision, 67.3% recall, and 72.5% F-measure, while the CNNs-based segmentation method achieved 94.3% precision, 49.5% recall, and 65.0% F-measure. The proposed method achieves higher recall and F-measure than the CNNs-based method, while the CNNs-based segmentation method achieves better precision. From the results, the proposed method reduces the number of missed potted flower plants compared to the CNNs-based segmentation method.

The reason for the 78.5% precision of the proposed method may be due to the low accuracy of segmentation in the dense areas of potted flowers. The images of unsuccessful segmentation are shown in Fig.4. From the left, the input image, the pre-processed image, and the image after the second Hough transform. The reason for the low segmentation accuracy is that the background area could not be detected due to the distortion from the center to the edges of the image. Recall is also considered to be limited to 67.5% due to the low accuracy of image segmentation as well as precision. Therefore, if the segmentation accuracy of the individual potted flower area can be improved by correcting the distortion of the image and adjusting the parameters of the Hough transform, the overall accuracy of the potted flower area segmentation method can also be improved.

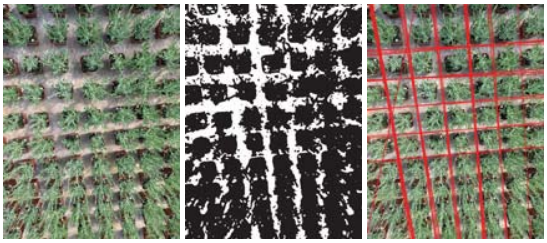


Figure 4: Unsuccessful Segmentation((a): Input image, (b): Pre-processed image, (c): Image after line Hough transform)

5 Conclusion

In this paper, we proposed a method to improve the accuracy of plant disease detection by image processing using the Hough transform. The results of trimming individual potted plant areas in a grid pattern feature using the Hough transform from images of the entire farm showed that one potted plant was trimmed with 72.5% F-measure, which is about 7 points higher than that of the segmentation method using CNNs.

As the results, proposal method achieved high accuracy in the segmentation with 136 images than R-CNN. And the system is capable of segmenting individual potted flowers as long as the target object is green and arranged in a grid pattern.

REFERENCES

[1] J. C. Zadoks: Crop Production and Crop Protection: Estimated losses in major food and cash crops, Agricultural Systems, Vol.51, No.4, pp.493–495 (1996).

[2] E. Fujita, H. Uga, S. Kagiwada, and H. Iyatomi: A Practical Plant Diagnosis System for Field Leaf Images and Feature Visualization, International Journal of Engineering & Technology, Vol.7, No.4.11, pp.49–54 (2018).

[3] K. Suwa, Q. H. Cap, R. Kotani, H. Uga, S. Kagiwada, and H. Iyatomi: A Comparable Study: Intrinsic Difficulties of Practical Plant Diagnosis from Wide-Angle Images, Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), pp. 5195–5201 (2019).

[4] Q. Yao, Z. Guan, Y. Zhou, J. Tang, Y. Hu, and B. Yang: Application of Support Vector Machine for Detecting Rice Diseases Using Shape and Color Texture Features, Proceedings of the 2009 International Conference on Engineering Computation, pp.79–83 (2009).

[5] T. Jadhav, N. Chavan, S. Jadhav, and V. Dubhele: A Review on Plant Disease Detection using Image Processing: International Research Journal of Engineering and Technology (IRJET), Vol.6, No.2, pp.2526–2530 (2019).

[6] Y. Kawasaki: Leaves Defection and Recognition Method for Automated Plant Disease Diagnosis, Bulletin of Graduate Science and Engineering, Engineering Studies of Hosei University, Vol.58, pp.1–4 (2017).(*in Japanese*)

[7] D. H. Ballard: Generalizing the Hough Transform to Detect Arbitrary Shapes, Pattern Recognition, Vol.13, No.2, pp.111–122 (1981).

[8] N. Otsu: A Threshold Selection Method from Gray-Level Histograms, IEEE Transactions on Systems, Man, and Cybernetics, Vol.9, No.1, pp.62–66 (1979).

[9] M. Ester, H.-P.Kriegel, J. Sander, and X. Xu: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD'96), pp.226–231 (1996).

A Feature Data Distribution Scheme for Person Tracking Systems with Multiple Cameras

Satoru Matsumoto* and Tomoki Yoshihisa*

*Cybermedia Center, Osaka University, Japan
{smatsumoto, yoshihisa}@cmc.osaka-u.ac.jp

Abstract-Recently, public cameras are widely used and are installed in various places. These multiple cameras can be used for tracking lost children or criminals. For person tracking, most systems send feature data such as feature values or person images to a server. The server compares the data and judges whether they are the same person. Artificial intelligence and numerical analyses can be used for the comparison. However, in this conventional scheme, the data amount that is sent to the server increases as the number of cameras increases. Hence, in this research, we propose a scheme for distributing the computational loads of the server arose in the conventional scheme. We propose two methods to determine the timing for camera devices to send feature data to other cameras. We evaluate these proposed methods and compare their performances. The simulation results show that the average traffic for each camera device can be reduced significantly compared to the average traffic under a conventional scheme.

Keywords: public cameras, feature data, processing servers, peer-to-peer

1 INTRODUCTION

Due to the recent trend of Society 5.0 and Smart city, the development of comfortable cities using IT technology has attracted great attention. Understanding how and when people move through the city can be useful in solving various social issues, such as marketing and research on human flow. Therefore, obtaining the travelling routes of people moving around the city contributes to the comfortable cities. If it is possible to track people using many security cameras installed in cities, we can track many people widely.

Some methods to detect a person in multiple images obtained from multiple cameras that do not share the same field of view have been proposed. These methods track the people by mapping same person across multiple cameras. These studies use deep learning or deep distance learning [1-4] to obtain feature representations with high discriminative power.

However, when identifying the same person from multiple images obtained from multiple cameras, a large amount of communication bandwidth is required if all the images are transmitted to the server. In addition, if all the information obtained from the video is sent to the server and the person re-identification process is performed on the server, the load on the server increases as the number of cameras and persons tracked increases. Even in the case of using cloud video recording services, the load on the recording server increases.

Hence, in this paper, we propose a person tracking method that does not concentrate the load on the server. In the proposed method, a camera network is built by multiple camera devices that can communicate with each other. When a person is captured in a camera's field of view, the feature values of the person image are calculated. The camera device then transmits the calculated feature values to the camera devices where the person is going to be captured next. The camera device that receives the feature values compares the feature value of each captured people with those received before and judges whether the captured person is captured before by other camera devices. Since each camera device performs the process of person re-identification using the model in the camera, our proposed method has a large possibility to suppress the concentration of the load on the server as the number of cameras and persons increases. Furthermore, to evaluate how much communication is generated when people are tracked using our proposed method, we create a simulator. We compare the average communication traffic of the server under a conventional method and that of our proposed method and confirm that our proposed method can distribute the load.

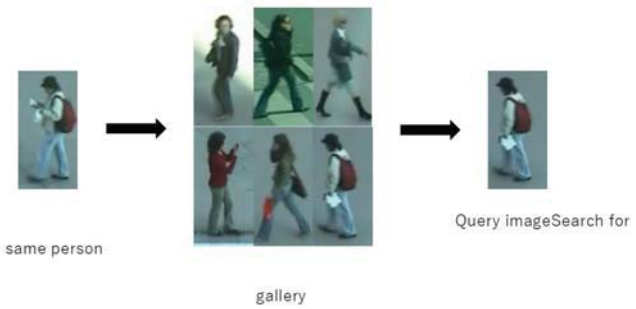


Fig 1. Overview of person re-identification

The organization of the paper is as follows. Section 2 inscribes existing research on person re-identification, a problem deeply related to this research. Section 3 inscribes the proposed method, and the evaluation results are shown in Section 4. Finally, we conclude the paper in Section 5.

2 RELATED WORK

Person re-identification is the problem of identifying the same person from images of people captured by multiple cameras that do not share the same field of view. Given a query image, the person re-identification system searches for a person identical to the query image in the gallery images (see Figure 1). Many studies improved the accuracy of person re-identification. Some of them use distance learning such as triplet loss, etc. [5-10].

Person re-identification is expected to have a wide range of applications in computer vision, such as surveillance,

behavior analysis, and person tracking, but on the other hand, it has a major problem due to gaps between cameras. When using multiple person images captured by multiple cameras that do not share the same field of view to perform person re-identification, the inter-camera gaps are unavoidable due to the nature that the person images used were captured by different cameras [6].

Therefore, person re-identification systems that can give a higher accuracy even when the influences of these varieties are large.

Many research has been done to overcome these effects. In [5], an adversarial network is used to obtain a more accurate feature representation that eliminates gaps between cameras as much as possible. The method proposed in [7] uses StarGAN to transform the styles of people in images. The method transforms the images of the people captured by a camera device to the images that consider the shooting conditions (background, lighting, etc.) of other cameras, then it uses these images as training data to reduce the influence of gaps between cameras. Also, there is a study that investigate how the variety of viewpoints affects the accuracy of person re-identification, as in [6].

Although many studies have been conducted to reduce the influence of above differences in conditions between cameras, the several problems still exist.

For the case where the camera images overlap, [8] performs partial figure re-identification using local features. To solve the problem of cloth changes, [9] systematically investigates how the accuracy of re-identification of existing models changes when clothes changes by generating pedestrian images with other clothes. In [10], a method that person re-identification with removing the external information of clothes and focuses on body shape information is proposed.

However, the systems that adopt these existing methods need to collect all camera images to a computational server. This causes a large communication and processing loads on the server. Even in the conventional method under that the cameras send only the feature values of the identified people to the server, the loads concentrate on the server. We aim to relief this loads for person re-identification in the paper.

3 PROPOSED METHOD

In this section, we first provide an overview of the proposed person tracking method. After that, we explain the detail.

3.1 Summary

Consider the case of a person tracking using surveillance cameras deployed throughout a city or a facility, using a conventional person re-identification method as described in Section 2. In this case, the method of transferring the images captured by the cameras to a server via a computer network and processing the person re-identification on the server requires a large communication traffic for the transmission of the images. Moreover, the methods in which information obtained from the video is transferred to the server and the person re-identification process is performed on the server increases the load on the server as the number of cameras or persons increases. To solve these problems, we propose a person tracking scheme in which features are transmitted among cameras. In our proposed method, a camera network is built using multiple camera devices that can communicate with each other, and the travelling paths of people in the target area are tracked by repeatedly sending and receiving feature values between camera devices and re-identifying people.

If re-identification fails, the system cannot track the person. Thus, the tracking performance can deteriorate compared with the system that a server manages all the cameras. However, our proposed system can distribute the communication and processing loads arose on the server in the above system.

3.2 Tracking Method

In this section, we describe the process flow of person tracking using camera device network. The proposed method is based on the following four assumptions.

- All camera devices that are connected to the camera device network can communicate with each other and send feature values.
- All camera devices have a neural network model that calculate the feature values of a captured person. The input of the model is him/her image. Each camera device gets the images from their connected cameras.
- The locations and the angles of the cameras are fixed, and the positioning of all cameras is assumed to be known in advance.
- All camera devices can estimate the direction of movement of a person using the coordinate and the interframe information.

Under the above assumptions, the camera devices connected to the computer network track the travelling path of a person by repeatedly transmitting feature values and judging whether the person is the same person. The following is an overview of the process flow when a person is successfully tracked between Camera A and Camera B.

1. Camera A captures a new person X.
2. Camera A detects a person, acquires a person image, and calculates the feature of the person X using a neural network model.
3. The destination camera device is determined by the destination determination method (detailed in 3.4) and the feature X is sent.
4. Camera B adds the feature X to the gallery.
5. A person moves and is captured by Camera B.
6. Camera B computes the feature values and compares them with the feature values X in the gallery to determine that they are the same person or not.
7. The fact that the person captured by Camera A was also captured by Camera B indicates that the person moved from A to B

.Figure 2 shows an image of a person re-identification process. In Figure 2, there are two separate images of people on the left side, and they are input to the same neural network model (NN model). The distance between the output features is calculated. The distance is between the features is used to judge whether the persons in the images are the same person or not.

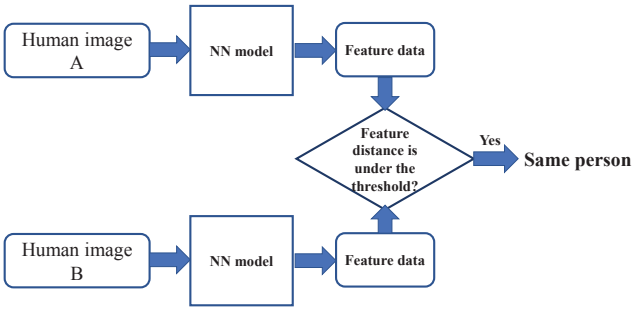


Fig 2. An image of a person re-identification process

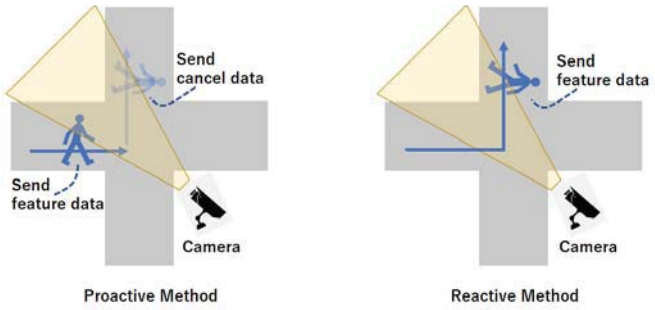


Fig 3. Time to start communication under our proposed method

3.3 Processes for Each Camera

The flow of processes executed by each camera device is shown below.

1. A person is captured by the camera.
2. Obtain bounding boxes and calculate features with NN models.
3. Person identification by comparing the calculated features with those in the gallery.
4. If these match, go to 7.
5. If these do not match, the feature values are sent to a camera device that is determined using the destination determination method (see 3.4) because the person is a newly detected person. The received camera device adds the feature values to the gallery.
6. Return to 1.
7. Notifies the server that a person has been detected.
8. The features of the person are again sent to the camera device determined using the destination determination method. The received camera device adds the feature values to the gallery.
9. Return to 1.

3.4 Destination Determination Methods

In this section, we describe a method for determining the destination camera device for transmitting feature values to another camera. In order to track a person, it is considered that when a person is captured by a camera, the feature values should be transmitted to its neighboring cameras. This is because the cameras neighboring to the camera device that a person captured is likely to be captured in the next. However, in the method where the feature values are transmitted only to the neighboring cameras, there is a possibility that the tracking of a person fails if the neighboring cameras fail to detect the person. One of the solutions for avoiding the failures is transmitting the feature values to further neighboring camera devices (the neighboring camera devices of the neighboring camera devices, etc.). Therefore, in the proposed method, we introduce a parameter N that indicates the number of the communication hops from the source camera device to send the feature values.

As described in section 3.2, camera devices can predict the direction of moving people, and therefore, it is possible to limit the transmission destinations by using the direction. That is, the direction of movement can be used to limit the transmission destination. The transmitted features are deleted after a certain time has elapsed, preventing features that are not used for tracking from remaining in the gallery.

Based on the above approach, we propose two types of methods for determining the transmission destination. The image of each method is shown in Figure 3. The first one is to transmit feature values to all the neighboring cameras within N hops when a person is captured by a camera device. The transmission timing is when the moving direction of the person is predicted. After the reception, the camera devices that are not likely to capture the person need to delete the feature values from the gallery (the proactive method). The other one does not transmit feature values when a person is captured by a camera device but transmits feature values to the camera devices that exist in the destination direction when the direction of the person is predicted (the reactive method).

3.4.1. Proactive Method

The flow of the proactive method is shown in Figure 4. The gray areas in the figure represent roads. The people walk on those areas. For simplicity, the roads are grid-shaped as shown in the figure, but the same process can be applied to roads that are not grid-shaped. The camera devices are assumed to be located at each intersection, and the locations of the camera devices are marked with the numbers (1 to 6). The parameter N is set to 2, which indicates how many cameras are to transmit the feature values to the next camera.

Procedure 1 shows how the features are transmitted when a person is detected by Camera 1. Camera 1 sends the feature values of the person to the surrounding N (= 2) camera devices when it detects a person. (Cameras with red numbers are the those hold the feature values.) In Procedure 2, the person moves from the area that Camera 1 shoots to the area that Camera 2 shoots. Camera 1 judges that Camera 2 is the camera that may capture the person in the next based on its direction.In Procedure 3, Camera 4 and Camera 6 are notified to remove the feature values from the gallery. This avoids the cameras that are unlikely to capture the person from continuing to have the feature values and reduces the number of candidates for the person re-identification. Figure 5 shows the processes for each camera device in the proactive method. In the proactive method, when a new person is captured in the field of view of a camera, it calculates the feature values of the person. Then, the camera device determines whether the captured person is the same person that other camera devices capture before, by calculating the distance among feature values. When it finds the same person, it sends only the information that the person was captured to the server. In the

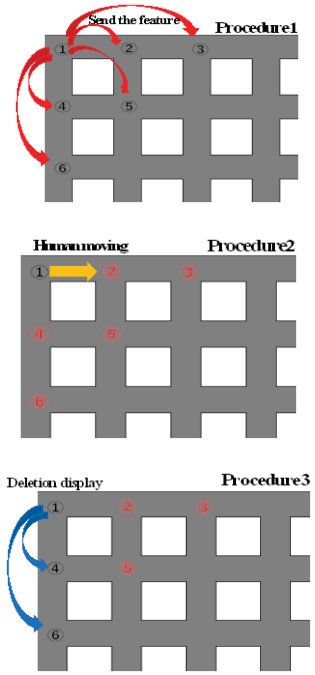


Fig 4. The flow of the proactive method

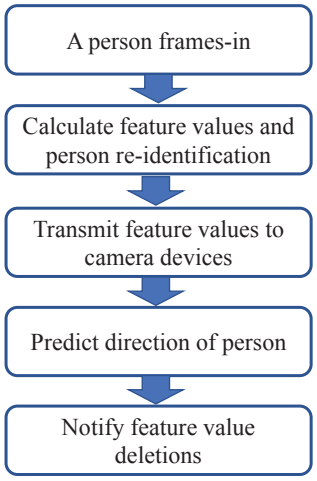


Fig 5. Processes for each camera device in the proactive method

proactive method, after that, the camera device transmits the feature value of the captured person as well as own Camera ID to *N* (at maximum) neighboring camera devices. In this case, the system can avoid duplicate transmissions because it is possible to find the camera devices to which the feature has been transmitted in the past from the list of camera IDs. If the same person is not found, it is assumed that the person is a new person and the feature is sent to all camera devices to *N* (at maximum) neighbors. If the direction of the person is predictable from the direction and the location information at the time of frame-out, the number of galleries for person re-identification can be reduced by notifying the camera devices to delete the features stored that is not likely to capture the person.

3.4.2. Reactive Method

The flow of the reactive method is shown in Figure 6. The road and the camera devices deployment are the same as the example for the proactive method in the previous subsection. Unlike the proactive method, the reactive method starts transmitting features after the moving direction of the person is found.

Procedure 1 shows the movement of the person from the area that Camera 1 shoots to that of Camera 2. In Procedure 2, the feature values are transmitted only to the camera device that exist in the direction of the person when Camera 1 detects it. We assume that the direction is predictable based on the travelling path of the person in the camera's field of view, such as the trajectory of the person and the position at which the person frames out.

The reactive method has the advantage of reducing the amount of communication because each camera device predicts the direction in which a person is moving and transmits the feature values only to the camera devices that exist in the direction. On the other hand, if the direction of the person cannot be predicted correctly, the feature values are not transmitted to the camera devices in the direction of the person, thus the tracking fails. If the terrain is complex, or if it is considered difficult to correctly predict the direction of a person due to the positional relationship among cameras, the probability of tracking failures can be high.

Figure 7 shows the processes for each camera device in the reactive method. In the reactive method, as in the proactive method, each camera device calculates feature values and re-identifies people when a new person is captured in the field of view. However, the feature values are not transmitted immediately, but only to the *N* neighboring cameras in the direction of their movement after predicting them based on their trajectories.

4 EVALUATION

To evaluate the amount of communication traffic generated when tracking a person under our proposed method, we created a simulator and measured the performances.

4.1 Simulation Specifications

To systematically evaluate the performance of our proposed methods, we assume that the roads are grid-shape. A camera network is built with camera devices that can communicate with each other and are located at each intersection.

In order to simulate different map sizes, we use three different maps as a 4 × 5 grid with 20 cameras, a 5 × 6 grid with 30 cameras, and a 6 × 7 grid with 42 cameras. One section of the grid is fixed by 10 meters. The map becomes larger as the number of cameras increases.

4.1.1. Parameters

We change the following five parameters in the simulator.

- The parameter to determine the number of the camera devices that receive the feature values. When the value is *N*, the feature values are transmitted to *N* neighboring camera devices.
- The number of persons flowing into the tracking area per a second.

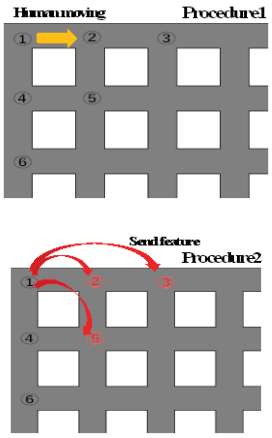


Fig 6. The flow of the reactive method

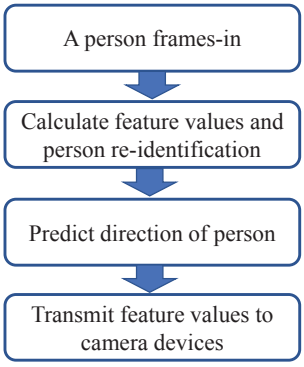


Fig 7. Processes for each camera device in the reactive method

- The number of the camera devices deployed in the tracking area. As described in the previous section, the number is one of 20, 30, or 42.
- Since a person entering an intersection is not always detected by the camera, the detection probability can be changed as a parameter ranging from 0.0 to 1.0. This value depends on perspective, lighting, and resolutions in real situations, but these conditions are various and thus we give the probability as a parameter.
- We set the communication bandwidth available for transmitting feature values. By setting the communication bandwidth, we can calculate the delay time for transmission.

4.1.2. Person Travelling Model

People walk at a speed of 1 meter per a second. Since one section of the grid is 10 meters long, the time between one camera capturing the person and the next is 10 seconds. A person enters the map at the upper left corner and exits at the lower left, the upper right, and the lower right corners. The number of people exiting from each exit is adjusted to be the same.

4.2 Evaluation Results

We get the results under the following situations.

4.2.1. Evaluation Items

The change in the communication traffic under the condition of different number of the camera devices and different people inflow.

The change in the tracking success rate in the proposed method changing the communication bandwidth. The tracking success rate is the rate that the number of the people that are tracked from the time to enter the tracking area to the time to exit divided by the number of the entered people. When the communication delays among the camera devices are all shorter than the one block travelling time of a person, the person is tracked in the tracking area.

Comparison of the average communication traffic of the server in a conventional method, in which the camera devices transmit the feature values to the server with that under our proposed method.

4.2.2. Communication Traffic for Transmitting Feature Values

We evaluated how the communication traffic changes when the number of cameras is changed to 20, 30, or 42, and when the number of people per second is changed to 1, 2, 3, or 4. In this evaluation, we assume that the communication traffic for one set of feature values is 1.648 [Kbit] assuming a 50-dimensional vector of float32 as the feature data. Also, three 16-bit regions are assumed to be reserved to record the ID identifying the person and the cameras that have passed through the area). This is an example setting. The detection probability for each camera device is set to 0.8.

The results are shown in Figure 8. The vertical axis represents the communication traffic for transmitting feature values. The unit is Kbps. The horizontal axis represents the number of people per second. From the results, it can be considered that there is a proportional relationship between the number of people and the communication traffic. The communication traffic ranges from 15 [Kbps] to 170 [Kbps] when the number of the camera devices is between 20 and 42 and the number of people per second is between 1 and 4.

We confirmed that there is a proportional relationship between the number of cameras and the amount of communication.

4.2.3. Tracking Success Rate

In order to track a person without tracking failures due to latency, it is necessary to provide more bandwidth than the amount of communication generated. If the amount of communication per second generated by tracking exceeds the bandwidth provided, the delay in sending features will increase as tracking continues, and the delay will diverge to infinity. If the bandwidth is not sufficient for the number of people travelling through the area, the tracking will fail due to delay.

4.2.4. Comparison of Communication Traffic

We simulated and compared the average communication traffic of the server (the amount of data received per unit of

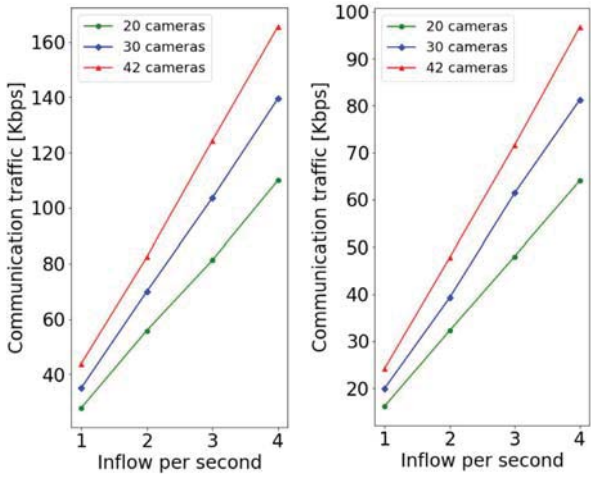


Fig 8. Communication traffic changing the number of camera devices and the people inflow (left: the proactive method, right: the reactive method)

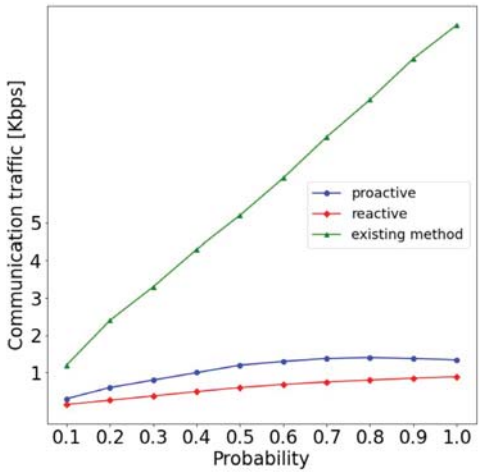


Fig 9. Comparison of communication traffic

time) and that under the proposed method. These traffic arises when all the feature values of the people captured by a camera device are transmitted to the server or other camera devices. The simulation results are shown in Figure 9. The horizontal axis is the person detection probability explained in Section 4.1.1.

The communication traffic for each camera device in our proposed method is reduced about one-twelfth at most compared to the average communication traffic under the conventional method, in which the server identifies the same person on the server. This indicates that the load that was concentrated on the server in the conventional method is distributed to each camera device under our proposed method.

5 CONCLUSION

A human tracking scheme in which each camera device sends the camera image to a server causes a large communication and processing loads on the server. This lengthens the delay for tracking and deteriorates the tracking success rate. Hence, in this research, we proposed a human tracking method in which each camera device transmits feature values of captured people among camera devices. We proposed two methods to determine the timing for camera devices to send feature data to other cameras. We developed a simulator for the evaluation and simulated the situation in that the number of the camera devices is 20, 30 or 42, and the tracking area is a grid-shape roads. Our simulation evaluation revealed that the average traffic per camera device under our proposed method can be reduced significantly compared to the average traffic on the server.

Future work includes the simulation of a more realistic environment and the evaluation on a real situation.

ACKNOWLEDGEMENT

This research was supported by a Grants-in-Aid for Scientific Research(C) numbered 21H03429, 20K11829, 20H00584 and by G-7 Scholarship Foundation.

REFERENCES

[1] D. Wu, S.-J. Zheng, X.-P. Zhang, C.-A. Yuan, F. Cheng, Y. Zhao, Y.-J. Lin, Z.-Q. Zhao, Y.-L. Jiang, and D.-S. Huang, "Deep learning-based methods for person re-identification: a comprehensive review," Neurocomputing, vol. 337, pp. 354-371, 2019.

[2] S. Liao and L. Shao, "Graph sampling based deep metric learning for generalizable person re-identification," arXiv:2104.01546, 11pages, 2021.

[3] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, and S. C. Hoi, "Deep learning for person re- identification: A survey and outlook," IEEE Transactions on Pattern Analysis and Machine Intelligence, 20pages, 2021.

[4] W. Wu, D. Tao, H. Li, Z. Yang, and J. Cheng, "Deep features for person re-identification on metric learning," Pattern Recognition, vol. 110, 12pages, 2021.

[5] Y. Li, R. Xue, M. Zhu, J. Xu, and Z. Xu, "Angular triplet loss-based camera network for reid," in 2021 International Joint Conference on Neural Networks (IJCNN), pp. 1-7, 2021.

[6] X. Sun and L. Zheng, "Dissecting person re-identification from the viewpoint of the viewpoint," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 608-617, 2019.

[7] Y. Lin, Y. Wu, C. Yan, M. Xu, and Y. Yang, "Unsupervised person re-identification via cross- camera similarity exploration," IEEE Transactions on Image Processing, vol. 29, pp. 5481- 5490, 2020.

[8] X. Zhang, Y. Yan, J.-H. Xue, Y. Hua, and H. Wang, "Semantic-aware occlusion-robust net- work for occluded person re-identification," IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 7, pp. 2764-2778, 2021.

[9] F. Wan, Y. Wu, X. Qian, Y. Chen, and Y. Fu, "When person re-identification meets change- ing clothes," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 830-831, 2020.

[10] X. Qian, W. Wang, L. Zhang, F. Zhu, Y. Fu, T. Xiang, Y.-G. Jiang, and X. Xue, "Long-term cloth-changing person re-identification," in Proceedings of the Asian Conference on Computer Vision, 17pages, 2020.

Session 3:
Multimedia
(Chair: Tomoki Yoshihisa)

Stationary Human Recognition by Multiple Vertically Arranged 2D-LiDARs

Takuya Watanabe*, Yuya Sawano*, Yoshiaki Terashima**, Ryozo Kiyohara †

*Graduate School of Kanagawa Institute of Technology, Japan
**Soka University, Japan
† Kanagawa Institute of Technology, Japan

Abstract - Several manufacturers have started the development of automated vehicle technologies which can be adopted for various automated robots or small carts. Human recognition technologies is one such kind of automated vehicle technology. In several cases, stationary human recognition is realized using point-cloud data from a three-dimensional laser imaging detection and ranging (3D-LiDAR). However, it requires numerous computer resources, such as a CPU and memories. In this study, a new human recognition method using point-cloud data from multiple vertically arranged 2D-LiDARs is proposed, and its performance is evaluated.

Keywords: stationary human detection, LiDAR, multiple vertically arranged 2D-LiDARs, point-cloud, auto-encoder

1 INTRODUCTION

The Kanagawa Institute of Technology is promoting a research project which is called the "KAIT Mobility Research Campus Project." The purpose of this project is to develop technologies for operating small autonomous robot carts (SARC) which can move freely on campus by integrating several technologies, such as AI, sensing and communication technologies etc.

The SARC should have certain sensors for the recognition of the outside and inside status. The camera and three-dimensional laser imaging detection and ranging (3D-LiDAR) are used for obstacle or human detection which require safety during operations and simultaneous localization and mapping (SLAM) technologies.

In this project, it is assumed that the SARCs are applied to the delivery robots for the documents and luggage in the daytime, to patrol the robots at midnight and to gather the fallen leaves in the morning.

For this purpose, the SARCs always estimate their own position using the SLAM technologies which enables sensing 360° around them via the 3D-LidARs. SLAM saves the position information and sensing data around it as the map data. Moreover, it always compares its sensing data and map data where it has visited to estimate its position.

The SARCs require the following functions for these purposes.

- The accuracy of the localization must satisfy that the SARC can be moved according to the plan.
- The SARC must avoid collision with the obstacles and humans.
- The SARC must keep a certain distance, for example, approximately 2 m, from a human to avoid issues.

To satisfy these requirements, there are following issues.

- The detection of humans
- The accuracy of localization

Till date, several human detection methods have been proposed as these technologies are required for autonomous driving. However, a majority of these technologies assume that high-specification cameras are installed and the distance can be detected by multiple cameras or a stereo camera. In reality, it is difficult to sense the distance using at monocular camera. If the size of object is known, the distance at which the human or another object is present can be estimated.

However, it cannot always be estimated. Therefore, it is desirable to be able to detect humans and measure the distance using only LiDAR. The point-cloud data from the 3D-LiDAR are quite large and require extensive computer resources. Therefore, a human detection method is proposed by using parts of the point-cloud data from the 3D-LiDAR, that is, multiple vertically arranged 2D-LiDARs.

The issue of detecting stationary humans is the focus of this study as there already exists extensive research for the detection of walking humans. In Section 2, the proposed project and the features of the LiDAR are explained. Previous related studies and the proposed method are presented in Sections 3 and 4, respectively. In section 5, the experiment is explained and the evaluations are presented in section 6. The final section includes a brief summary.

2 KAIT MOBILITY RESEARCH CAMPUS

2.1 Objective of the Project

The "KAIT mobility research campus" is envisioned as a project in which autonomous small carts move freely by themselves in the campus and making life and learning rich and enjoyable.

The campus has high level infrastructure which have been created using communication and information processing technologies. In this campus, SARCs which have various

roles move about everyday by themselves (examples are shown in Figure 1). The students and teachers can use, watch, and be guarded using these SARCs. At the same time they can develop their own SARC and experiment with it.

2.2 SARC Architecture

The SARC architecture is hierarchical mobile intelligence architecture which is integrated using several technologies such as AI, sensing, recognizing, and communication technologies, as shown in Figure 2. According to the status around the SARC, it is controlled properly by hierarchical mobile intelligence. This architecture is common between SARCs which have multiple roles.

2.3 LiDAR

The LiDAR is the most important and essential sensor for a SARC. It is a kind of sensor which can find and measure the distance between objects using laser light. The objects are detected using scattered light and the distance between them is measured using the phase difference. Moreover, the materials are estimated using the attenuation rate. Therefore, the types of data obtained from a LiDAR are more than the types of data obtained from images captured by a camera. The high precision of the around status by a LiDAR can be used for terrain measurement. Recently, this technology has been applied for the localization of autonomous carts, detection of obstacles, collision avoidance for vehicles, autonomous driving etc.

However, in bad weather conditions, the obtainable distance is decreased or there are false detections. Especially, in the cases where the light reflection cannot be received, the data cannot be gathered, for example, if the glass does not reflect the light. Further, it is difficult to detect a mirror or aluminum sheet because it does not scatter the light [1].

There are two types of LiDARs: solid type and rotation type. The rotation type has a wide range of sensing which creates modules which irradiate the laser light rotating it 360°. However, it has low durability. The solid type has no physically moving parts. However, the detection range is narrower than the rotation type. Furthermore, LiDARs can be classified as 3D-LiDARs and 2D-LiDARs. A 3D-Lidar has multiple laser lights in which each light has a different height. The point-cloud data are shown in Figures 3 and 4. Figure 3 shows an example of the point-cloud data obtained from a 2D-LiDAR which can only gather a small amount of data. This makes it difficult to recognize any object as a

Figure 1. Image of examples of SARC

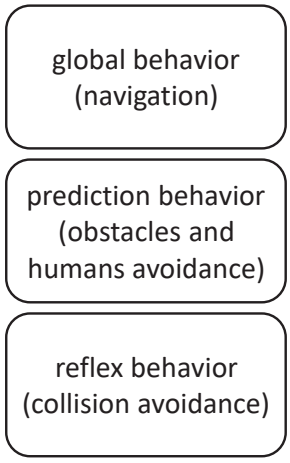


Figure 2 Hierarchical mobile intelligence architecture

Figure 3. Example of point-cloud data using a 2D-LiDAR

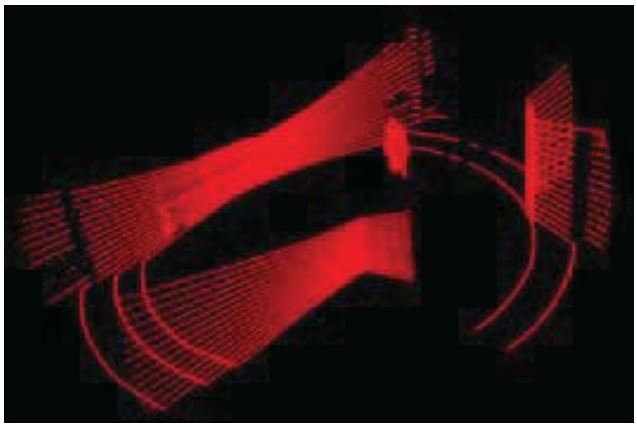


Figure 4. Example of point-cloud data using a 3D-LiDAR

human from solely its shape. However, the equipment cost for a 2D-LiDAR is low. Figure 4 shows an example of the point-cloud data obtained from a 3D-LiDAR that can gather a larger amount of data for purposes, such as autonomous vehicles which require a higher accuracy. The 2D-LiDAR is ReThe point-cloud data from a 2D-LiDAR does not require extensive computer resources.

Therefore, in this study, it is assumed that multiple 2D-LiDARs which are vertically arranged can be used for stationary human detection.

3 RELATED STUDIES

Human detection and tracking have been studied extensively in recent years [2-5]. RGB-cameras can detect humans by recognizing the image data [2-3]. Additionally, RGB-D cameras can recognize humans [4-5]. However, these cameras require certain brightness conditions. Therefore, they cannot be used outdoors at night.

There have been various studies for the detection of humans using LiDAR [6-7]. Humans can be recognized through their shape which can be detected using a 2D-LiDAR and time series observation information [6]. However, the time series information of stationary humans cannot change every time. Otherwise, it can be assumed that the human is walking.

[7] is based on 2D-LiDARs which track two layers, waists, and knees of people, and can detect and follow multiple walking humans. In this study, the indoor environment and a walking human are assumed. Therefore, it does not recognize the human using the feature of human shape; however, it recognizes the moving objects are humans. The Kalman filter is used for the prediction of walking humans. The stationary humans are treated as background images and deleted.

There are a few studies [8-9] that have detected objects by sensor fusion technologies which use multiple different sensors, for example, a LiDAR and camera. The image processing technology can detect the objects from the image data, and the LiDAR technology can measure the distance by matching both sets of data.

These kinds of approaches might be able to detect objects or humans with high accuracy. However, the cost is high. Therefore, their installation in autonomous vehicles might be possible. In contrast, the cost is too high for installation in autonomous carts.

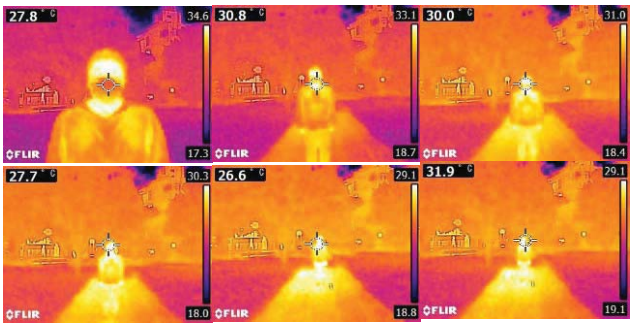


Figure 5 Example data of pyroelectric infrared sensor

A pyroelectric infrared sensor is one of the sensors which can detect people [10-11]. However, it assumes that humans walk. Examples of stationary humans are shown in Figure 5 in various situations and distances. This type of method might be able to detect humans depending on the situation. In the case of authorized people having smartphones and an application for authentication is installed in the phone, humans can be easily detected by bluetooth low energy (BLE) technology [11]. However, this can only be for authorized people with smartphones.

A human detection method using 2D point-cloud data and a deep learning technology is proposed [12]. This approach is applied in which human-crafted features is inputted to the one-class classification model to identify whether it is a human or not. However, this technology also assumed that humans walk.

In this study, a stationary human recognition method is proposed. A method for detecting humans using 2D point-cloud data with LiDAR [13] was proposed in a previous study by the authors. The results of the study showed that high accuracy could be obtained under certain conditions.

4 PROPOSED METHOD

In a previous study [13], the authors focused on the feature of the arm and waist to detect humans. The results showed 93% accuracy for human detection. However, as people change their clothes depending on the season their shapes can further change depending on the presence or absence of layers of clothes. Moreover, there are countless types of body shapes. Therefore, it is necessary to gather more data regarding the different body shapes. This will imply that the data which must be learned will increase along with the calculation cost. As a result, the ambiguity might increase, and the accuracy might decrease.

It is assumed that the shape of the human shin which is not dependent on the season can obtain high accuracy with the same volume of learning point-cloud data .

Hence, a human detection method with multiple point-cloud data from multiple vertically arranged 2D-LiDARs is proposed, as shown in Figure 6. If both sets of data are highly related, a high accuracy of detection of humans is expected. However, it must be implemented carefully.

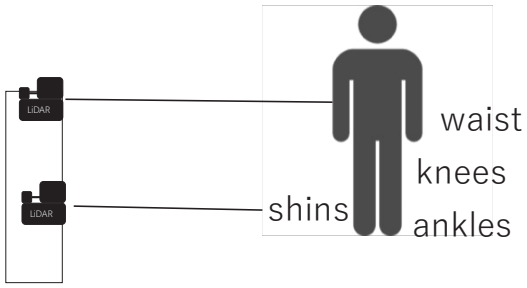


Figure 6. Multiple vertical arranged 2D-LiDARs s

Result of estimation \ correct	human	not human
human	A	B
not human	C	D

Figure 7. Accuracy of human detection

There are certain relations between the actual correct answer and the estimation results as shown in Figure 7. Generally, ratios A and D should be high as depicted in Figure 7. There can be two implementation methods. The first method is that both the data sets are learning at the same time. The second method is that both the data sets are leaning independently. The latter method is more straightforward and requires less calculation cost. However, A might be high, and D might be low. Therefore, it is suggested that the former method be adopted

5 EXPERIMENT & EVALUATION

Figure 8 An example of similar object outside campus.

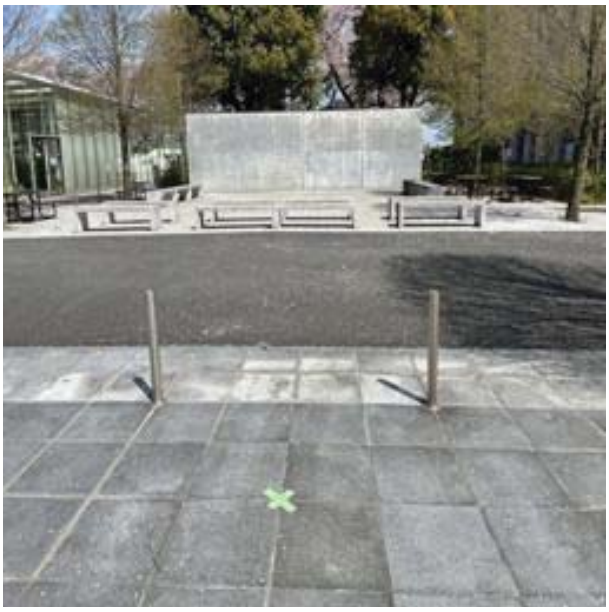


Figure 9 An example of similar object inside campus

5.1 Basic Experiment

The experiments were carried out to simply estimate the accuracy of human detection through the proposed method. There are similar objects to the human shin, for instance, various objects which have two feet were further utilized in the experiment. An example of an object with two feet seen outside of campus is shown in Figure 8. Figure 9 shows a similar object inside the campus. The data collected from these objects using 2D-LiDAR were used as incorrect data. The data collected from humans using 2D-LiDAR were considered as the correct data. A large volume of data is required for learning. The point-cloud data were gathered covering 360° of the LiDAR. However, only half of the data can be deleted as the data for learning might be symmetric in shape.

First, the essential information is gathered through certain experiments. The experiment environment was prepared for gathering the correct data, i.e., the shin information using 2D-LiDAR, as depicted in Figure 10. The experimental collaborators were asked to stand on the bench. The data for each measure of distance, i.e., 1, 2, 3 and 4 m were collected. Moreover, the experimental collaborators had to stand in the front, 45° diagonal to the right and left, and towards the back. Subsequently, the incorrect data were also gathered, as shown in Figure 11.

For confirmation of the proposed approach, experiments were carried out on a small scale. Both the correct and incorrect data were obtained from 20 people and 20 objects. The total amount of data was quite small. However, it could be used to confirm the possibility.

5.2 Evaluation

Figure 10 Experiment environment for gathering correct data

Figure 11 Example of incorrect data

The test data are classified to two classes using the support vector machine (SVM) [14]. The data regarding a symmetric object or human were evaluated by using half of point-cloud data. The detection ratio was approximately 41 % which is not excellent. Moreover, the data were classified by different distances, i.e., 3 or 4 m was classified as correct data and 1 or 2 m was considered as incorrect data for detection of only shin data. The detection ratio was about 12 % which is much lower than the previous one.

Table 1. Detection ratio for each angle

0°	45°	90°	135°
0	0	50%	0

Table 1 show the detection ratio for each angle. The angle to the LiDAR was changed to 0°, 45°, 90°, 135°. In the case of 90°, the detection ratio was 50%. Because, in the case of 90°, it is difficult to detect two shins. The data shows up as one pole for two shins. These results show it is difficult to detect humans using only the shin.

6 DISCUSSION AND CONCLUSION

The results of this study show that humans can be detected by the data of more than two 2D-LiDARs. Therefore, the human detection function should peep into the point-cloud data of 3D-LiDAR for SLAM. It should be able to select suitable point-cloud data from all data. Future studies will utilize two kinds of more than two data of 2D-LiDARs. The first one will be 2D-LiDARs by different heights and the second one will be 2D-LiDARs on more than two robot carts.

REFERENCES

[1] T Nishida, F Ohkawa, H Miyagawa, M Obata, “Development of a Sensor System for an Outdoor Service Robot,” INTECH Open Access Publisher, 2008.

[2] C. Premebida, J. Carreira, J. Batista, U. Nunes, “Pedestrian detection combining RGB and dense LIDAR data,” In Proceedings of the 2014 IEEE/RSJ international conference on intelligent robots and systems (IROS) ,pp. 4112–4117,2014.

[3] A. González, G. Villalonga, J. Xu, D. Vázquez, J. Amores, A.M. López, “Multiview random forest of local experts combining RGB and LIDAR data for pedestrian detection,” In Proceedings of the 2015 IEEE intelligent vehicles symposium (IV), pp. 356–361, 2015.

[4] T. Linder, S. Breuers, B. Leibe, K.O. Arras, “On multi-modal people tracking from mobile platforms in very crowded and dynamic environments,” In Proceedings of the 2016 IEEE international conference on robotics and automation (ICRA) ,pp. 5512–5519, 2016.

[5] M. Munaro, E. Menegatti, “Fast RGB-D people tracking for service robots,” Autonomous Robots, 37, pp. 227–242,2014.

[6] L. Sun, Z. Yan, S.M. Mellado, M. Hanheide, T. Duckett, “3DOF pedestrian trajectory prediction learned from long-term autonomous mobile robot deployment data,” In proceedings of the 2018 IEEE international conference on robotics and automation (ICRA), 2018.

[7] M. Hashimoto, T. Konda, Z Bai., K. Takahashi, "Laser-based tracking of randomly moving people in crowded environments", 2010 IEEE International

Conference on Automation and Logistics, pp. 31-36, 2010.
[8] M. Liang, B. Yang, Y. Chen, R. Hu, R. Urtasun, “Multi-Task Multi-Sensor Fusion for 3D Object Detection,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7345-7353, 2019.
[9] J. Kocić, N. Jovičić and V. Drndarević, "Sensors and Sensor Fusion in Autonomous Vehicles," 2018 26th Telecommunications Forum (TELFOR), pp. 420-425, 2018.
[10] J. Yun, S.-S. Lee, “Human Movement Detection and Identification Using Pyroelectric Infrared Sensors,” Sensors Vo.14, No.5 pp. 8057-8081, 2014.
[11] X. Luo, B. Shen, X. Guo, G. Luo, G. Wang, "Human tracking using ceiling pyroelectric infrared sensors," 2009 IEEE International Conference on Control and Automation, pp. 1716-1721, 2009.
[12] Y. Kohara and M. Nakazawa, "Human Tracking of Single Laser Range Finder Using Features Extracted by Deep Learning," 2019 Twelfth International Conference on Mobile Computing and Ubiquitous Network (ICMU), pp. 1-5, 2019.
[13] Y. Nagai, Y. Sawano, Y. Terashima, T. Suzuki and R. Kiyohara, "A Method for Detecting Human by 2D-LiDAR," 2022 IEEE International Conference on Consumer Electronics (ICCE), 2022. doi: 10.1109/ICCE53296.2022.9730524.
[14] SVM_pcl, https://github.com/bellonemauro/SVM_pcl_demo, <accessed 03/Jun/2022>
[15] Z. J. Chong, B. Qin, T. Bandyopadhyay, M. H. Ang, E. Frazzoli, D. Rus, “Mapping with synthetic 2D LIDAR in 3D urban environment,” IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4715-4720, doi: 10.1109/IROS.2013.6697035, 2013

Development and Evaluation of a Broadcaster Support Method using MR Stamp in 360-degree Internet Live Broadcasting

Yoshia Saito* and Kei Sato*

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

Abstract - In this study, we investigate the use of MR (Mixed Reality) stamp which supports broadcasters to reduce communication errors from viewers to a broadcaster in 360-degree Internet live broadcasting. There is a problem that the broadcaster is unable to grasp the viewer's POV compared to the conventional broadcasting method. We have confirmed that this problem could be reduced by combining an equirectangular video and 2D stamp which presents the viewer's interests in a simple image on the video. On the other hand, the 2D stamp system remains three issues. The issues are (1) the stamp cannot be fixed on the target object, (2) the broadcaster must check a PC display to see the stamp, and (3) it is difficult to understand a position of the stamp in the real space.

In this study, we propose MR stamp which can be displayed and fixed on the real space, which enables the broadcaster to check a holographic stamp on the real space through an MR device. To realize our proposal, we implemented the proposed system using Microsoft's HoloLens 2 which is an HMD-type MR device. It can show holograms in the real space by recognizing real space. We also evaluated the effectiveness of the MR stamp compared with the 2D stamp and found that the MR stamp with the spatial audio solved the three issues of the 2D stamp.

Keywords: 360-degree Internet live broadcasting, Mixed Reality, MR stamp, Broadcaster support method

1 INTRODUCTION

YouTube started a 360-degree Internet live broadcasting service from 2016 and it enables anyone to easily use the 360-degree Internet live broadcasting service now. The 360-degree Internet live broadcasting is a service which combines the Internet live broadcasting with 360-degree videos using an omnidirectional camera. In the 360-degree Internet live broadcasting, a broadcaster can provide a 360-degree video to viewers in real time without care about the view angle of the camera. The viewers can change the point of view (POV) according to their own interests and communicate with the broadcaster using text chat.

The 360-degree Internet broadcasting, however, has a lot of new problems. One of the problems is that the broadcaster cannot be aware of the viewers' POV. In the conventional Internet live broadcasting, it uses a web camera which has a single lens and the single lens definitely shows the viewers' POV. The broadcaster can see what they are watching by direction of the lens. On the other hand, in the 360-degree Internet live broadcasting, it uses an omnidirectional camera which has a wide-angle lens or multiple lens. It prevents the broadcaster to see what the viewers are watching by direction of the lens.

There are many studies about the role of gaze information in the remote communication [1][2]. In the studies, it concludes that the communicatee's gaze information indicates the target of interest or center of the topic. The gaze information in the remote communication is similar to the viewers' POV in the 360-degree Internet live broadcasting. The broadcaster sometimes cannot understand the context of the viewers' comments and it causes communication errors between the broadcaster and the viewers.

To solve the problem of the communication errors, we have proposed stamp functions in the 360-degree Internet live broadcasting [3]. In this study, the stamp functions help the broadcaster to understand what the viewers talk about and find the object. However, it remains three issues in the proposed system, which are (1) the stamp cannot be fixed on the target object, (2) the broadcaster must check a PC display to see the stamp, and (3) it is difficult to understand a position of the stamp in the real space. These issues should be solved to realize smooth communication between the broadcaster and the viewers.

In this study, we propose MR stamp which is a new stamp function using MR and can be displayed and fixed on the real space. It enables the broadcaster to check a holographic stamp on the real space through an MR device. The MR stamp can solve the issues of the previous stamp function.

The contributions of this paper are summarized as follows:

- We proposed a new broadcaster support method called MR stamp in 360-degree Internet live broadcasting.
- We developed and evaluated a prototype system of the MR stamp using HoloLens 2.
- We clarified that the MR stamp enabled the broadcaster to find a target object in a relatively short time and spatial audio could help the broadcaster to find the MR stamp.

The rest of this paper is organized as follows. Section 2 describes our previous work about stamp functions in the 360-degree Internet live broadcasting. Section 3 describes an overview of the proposed system and a use case of the MR stamp. Section 4 describes implementation of the proposed system. Section 5 describes a first evaluation

experiment to clarify the effects of the proposed system and reveal its problem. Section 6 describes a second evaluation experiment improving the proposed system. Section 7 summarizes this study.

2 PREVIOUS STUDY

We proposed have stamp functions in 360-degree Internet live broadcasting to support communication between the broadcaster and the viewers. In this section, we explain the stamp functions in the previous study and its known issues.

2.1 Stamp functions in 360-degree Internet live broadcasting

There are two stamps which are “Look” and “Go” stamp in the previous study as shown in Fig. 1. Figure 2 shows a user interface of the viewer in the previous study. The viewers can watch the 360-degree live video in spherical format and change the POV as they want. The viewers are also able to use Look and Go stamp by selecting the kind of stamp and clicking on the video. The stamps are shown in same place of the video on the user interface of the viewers and the broadcaster. Figure 3 shows a user interface of the broadcaster. The 360-degree video is displayed in equirectangular format. The broadcaster can check the stamps which are sent from the viewers without changing the POV.

We evaluated the previous stamp function and found their advantages. The stamp function improved easiness of communication between the broadcaster and the viewers compared with the case that the stamp was not used. The broadcaster could easily understand what the viewer talked about and found the object. Moreover, the stamp function increased the frequency of communication between the broadcaster and the viewers.

2.2 Known issues

Although the previous study has advantages, there are three issues as follows.

(1) The stamp cannot be fixed on the target object.
A stamp only has information of a direction from the omnidirectional camera at a particular time. Therefore, the stamp cannot be fixed on the target object and it causes positional shift of the stamp when the omnidirectional camera is moved.

(2) The broadcaster must check a PC display to see the stamp.
In the previous study, the broadcaster has to carry a laptop PC and check stamps from the viewers through the display. The check of the laptop PC display consumes time and prevents smooth communication between the broadcaster and the viewers. Besides that, it is dangerous to see the laptop PC while walking and it is inconvenient not to be able to use a hand which keeps the laptop PC.

(3) It is difficult to understand a position of the stamp in the real space.
The stamps are shown on the 360-degree video in equirectangular format. When a stamp is appeared on the video, the broadcaster must find a position of the stamp in the real space. The distortion of the 360-degree video in equirectangular format makes it difficult to find the direction of the stamp.

Figure 1: Stamps in the previous study

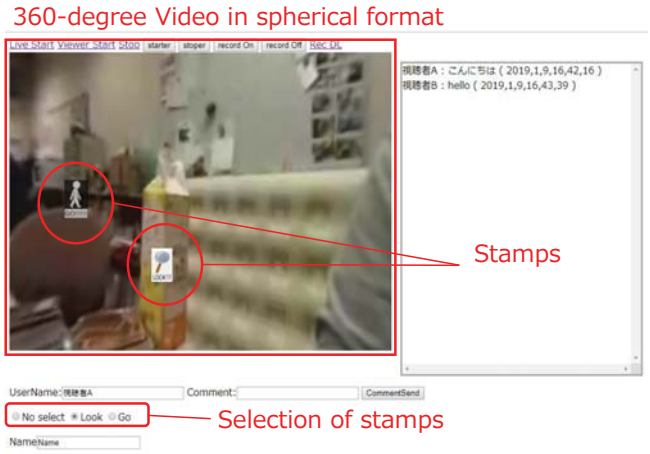


Figure 2: User Interface of viewer in the previous study

Figure 3: User Interface of broadcaster in the previous study

3 PROPOSED SYSTEM

The MR stamp aims to solve the issues of the previous study using MR and realize smooth communication between the broadcaster and the viewers.

3.1 Effectiveness of MR

MR is a technology which displays holograms of virtual objects in the real space and the users can interact with the holograms. Several researches show effectiveness of MR in remote communication between users.

Lee [4] developed a MR remote collaboration system which shared 360-degree live video. In this system, a hologram of remote user's hand is displayed on the real space through the MR device. The hand gestures by the hologram help to understand each other's focus and improve their communication. Johnson [5] studied the effect of MR guidance. An experiment was conducted to understand how providing explicit spatial information in collaborative MR environment. From the experiment result showed the MR guidance realized effective referencing through deixis.

From the related work, the reduction effect of communication errors can be also expected by introducing the MR technology for the stamp function in 360-degree Internet live broadcasting.

3.2 System Model

Figure 4 shows a model of the proposed system. The proposed system is included in the existed 360-degree Internet live broadcasting system. A broadcaster provides 360-degree live video to viewers using the 360-degree Internet live broadcasting system. The viewers can send a 2D stamp which is implemented in the previous study to the broadcaster. The proposed system receives 2D position information of the stamp and transforms it to 3D position information. The proposed system displays a MR stamp using the 3D position information and the broadcaster can check the MR stamp in the real space through a MR device.

The proposed system can solve three issues in the previous section. The first issue which is that "The stamp cannot be fixed on the target object" can be solved by fixing the stamp on the real space using MR. The second issue which is that "The broadcaster must check a PC display to see the stamp" can be solved by using a MR device which is a type of head-mounted display. The third issue which is that "It is difficult to understand a position of the stamp in the real space" can be solved by directly displaying the stamp on the target object in the real space using MR.

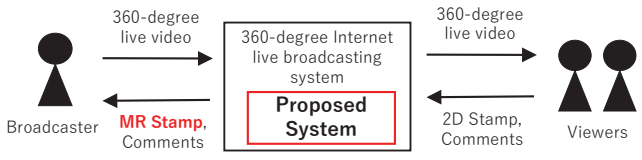


Figure 4: A model of the proposed system

3.3 Use Case

In this study, we suppose that an omnidirectional camera is fixed on an arbitrary position and 360-degree Internet live broadcasting is performed indoors. The reason is that the usage environment should be simple for the first step of this study. We also suppose it is used at the showroom and the exhibition.

Figure 5 shows a use case of the proposed system. The viewers can send a MR stamp to the broadcaster if they want to see a particular object in a showroom. The broadcaster can understand the request from the viewers easier than when only comments are used. Since 2000, online virtual events have been increasing because of COVID-19. In the online virtual events, it is said that more opportunity for real two-way communication between the broadcaster and viewers [6]. In online conferences, it is reported that it cannot have smooth relationship with each other between participants [7]. The MR stamp would realize smooth two-way communication between the broadcaster and the viewers in various online virtual events.

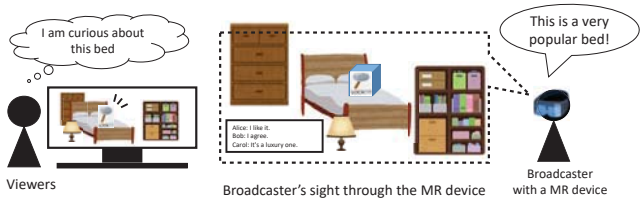


Figure 5: A use case of the proposed system

4 IMPLEMENTATION

We implemented a prototype system of the MR stamp using Hololens2. In this section, we describe the architecture of the prototype system and its main application.

4.1 System Architecture

The prototype of the proposed system is based on 360-degree Internet live broadcasting system of the previous study. Figure 6 shows the architecture of the prototype system. The red square shows new implementation in this study and the other parts are diverted from the previous study. A broadcaster can start 360-degree Internet live broadcasting using the client for broadcaster on a web browser. The 360-degree Internet live broadcasting server distributes it to viewers. The viewers can watch the 360-degree live video and send 2D stamp and comments in the same manner as the previous study. The stamp/comment server forwards it to all clients for viewers and a MR device of the broadcaster. We use Microsoft HoloLens 2 as the MR device. In the HoloLens2, the MR stamp application is running. The application presents the MR stamp and comments to the broadcaster. The broadcaster can check the MR stamp in the real space through the HoloLens 2.

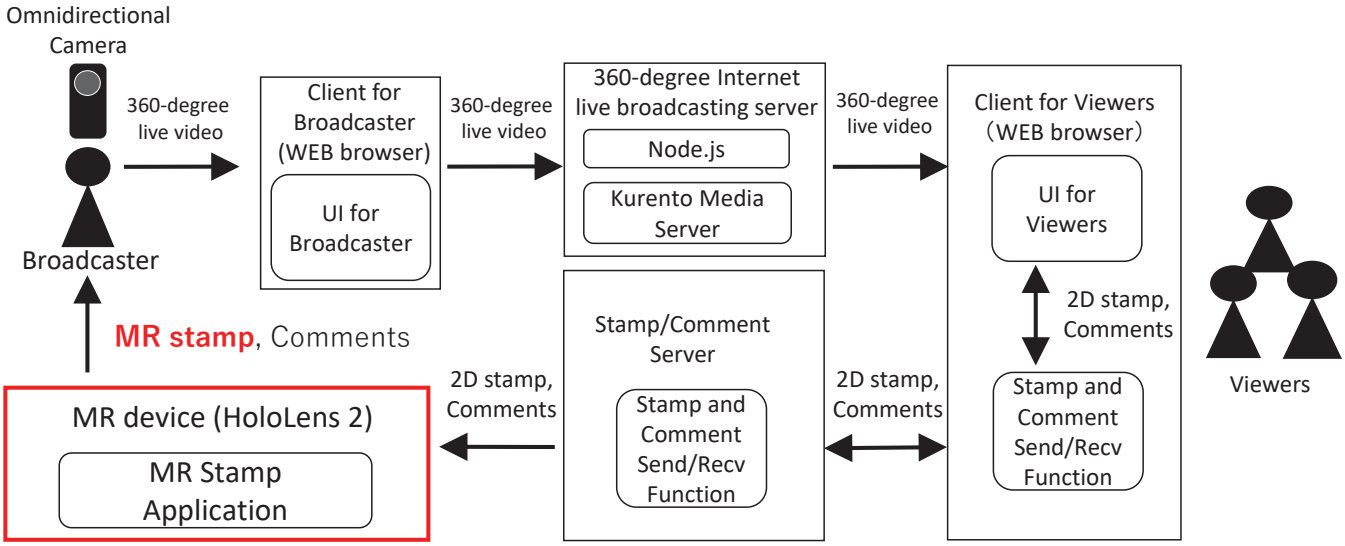


Figure 6: Architecture of the prototype system

4.2 MR stamp application

The MR stamp application receives 2D position information of the stamp and needs to transform it to 3D position information for the MR stamp. To realize the coordinate transformation, we use a Raycast function in Unity, which irradiates a 3D ray from an origin point to a target direction and detects intersecting collisions. The origin point is the coordinates of the omnidirectional camera. The target direction can be given by the coordinates of the stamp. The HoloLens 2 have a function of spatial mapping. The spatial mapping provides a detailed representation of real space surfaces in the environment around HoloLens 2. The raycast detects the intersecting collisions with the real space surfaces and returns the 3D coordinates. The MR stamp is displayed on the 3D coordinates.

Figure 7 shows the example of the MR stamp. The MR stamp is shown as a 3D square frame so that it is easy to check the target object from the broadcaster. The MR stamp disappear after the period of time (10 seconds in this implementation).

The comments from the viewers are displayed on a comment window as shown in Fig. 8. The comment window tracks the broadcaster's sight.

Figure 7: An example of the MR stamp

Figure 8: Comment window through HoloLens 2

5 FIRST EVALUATION

We conducted an experiment to evaluate effectiveness of the MR stamp using the implemented prototype system. The purpose of the evaluation is to check whether the three issues of the previous study can be solved or not by using the MR stamp.

5.1 Environment and Procedure

We compare the prototype system (MR stamp system) with the stamp system of the previous study (2D stamp system). In the experiment, A broadcaster performs 360-degree Internet live broadcasting and two viewers watch the broadcasting. In the room of the broadcaster, there are various objects. The viewers talk about an object in the broadcaster's room using the MR/2D stamp and comments. The broadcaster looks for the target object and communicates with the viewers. The evaluation items are as follows; (1) time to find the target object, (2) number of communication errors, and (3) subjective easiness to find the target object.

The experiments were conducted 4 times. There were one broadcaster and two viewers per time and the participants were students in our university. The broadcasting time was 30 minutes. Figure 9 shows the procedure of the experiment. At first, a viewer sends a stamp to the broadcaster. The target object and the viewer who performs the task are predetermined by the task instruction. The broadcaster looks for the target object referring to the stamp. After the target object is found, the broadcaster confirms whether the object is correct or not by speaking to the viewer. The viewer replies whether it is correct or not. If it is not correct, the broadcaster continues to find the target object. The search task with stamp is performed 2 times. At last, we ask the broadcaster how easy to find the target object in 5-point scale by a questionnaire. This procedure was performed for both the MR stamp system and the 2D stamp system. The order of the MR stamp and 2D stamp system was random to keep fairness.

Figure 10 shows the target objects used in the experiment. Target A and C are comprehensible ones to find because there is nothing around the object (hereafter, single object). Target B and D are mistakable ones to find because there are several similar objects around the object (hereafter, multiple objects). In addition to these target objects, there are several dummy objects in the broadcaster's room. Figure 11 shows the location of all objects in the broadcaster's room.

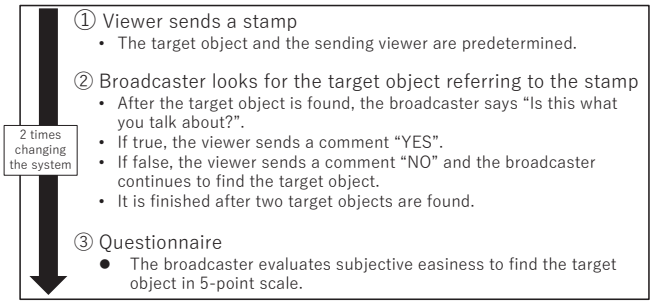


Figure 9: Procedure of the experiment

Target A : Black backpack

Target B : Blue labeled folder

Target C : Coffee maker

Target D : Stuffed white tiger on a cardboard box

Figure 10: Target objects

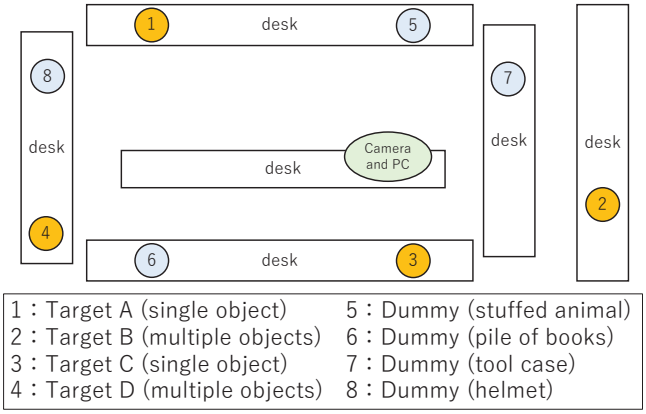


Figure 11: Location of all objects in the broadcaster's room

5.2 First Evaluation Results

Table 1 shows evaluation result of the first evaluation. There are the results of 4 broadcasters with 2D and MR stamp system. In terms of Target, "Single" means single object such as Target A and C and "Multiple" means multiple objects such as Target B and D. "Time to find" denotes time from when a stamp was displayed to when the broadcaster found the correct target object. "Errors" denotes number of times which the broadcaster indicated incorrect objects.

The average of time to find the target object in the 2D stamp system takes 11.25 seconds in case of the single object and 20.25 seconds in case of the multiple objects. The total of the average time in the 2D stamp system is 15.75 seconds. In the MR stamp system, it takes 18 seconds in case of the single object and 13.25 seconds in case of the multiple objects. The total of the average time in the MR stamp system is 15.625 seconds. The average of number of errors in the 2D stamp system shows 0.25 and 0.5 in the single and multiple objects respectively. The total of the average number of errors in the 2D stamp system is 0.375. The average of number of errors in the MR stamp system shows 0.25 and 0.25 in the single and multiple objects respectively. The total of the average number of errors in the MR stamp system is 0.375. We conducted a Student's t-test for the results of the total average and there is no significant difference both in the average time to find ($p = 0.98$) and the average number of errors ($p = 0.63$) between the 2D and MR stamp systems.

Table 2 shows the result of the questionnaire which is about subjective easiness to find the target object in 5-point scale. The average points of the 2D stamp system are 4.5 and 3.25 in the single and multiple objects respectively. The total average in the 2D stamp system is 3.85. The average points of the MR stamp system are 4.25 and 3.75 in the single and multiple objects respectively. The total average in the MR stamp system is 4. We conducted a Student's t-test for the results of the total average and there is no significant difference about the subjective easiness to find the target object ($p = 0.83$).

Table 1: Evaluation result of the first evaluation

Broadcaster	System	Target	Time to find (sec)	Errors (No. of times)
A	2D Stamp	Single	13	0
		Multiple	35	1
	MR Stamp	Single	8	0
		Multiple	16	1
B	2D Stamp	Single	15	1
		Multiple	24	1
	MR Stamp	Single	18	1
		Multiple	13	0
C	2D Stamp	Single	8	0
		Multiple	5	0
	MR Stamp	Single	19	0
		Multiple	9	0
D	2D Stamp	Single	9	0
		Multiple	17	0
	MR Stamp	Single	27	0
		Multiple	15	0
Average	2D Stamp	Single	11.25	0.25
		Multiple	20.25	0.5
		Total	**15.75**	**0.375**
	MR Stamp	Single	18	0.25
		Multiple	13.25	0.25
		Total	**15.625**	**0.25**

Table 2: Questionnaire result of the first evaluation

System	Target	Subjective easiness to find the target object (5-point scale)
2D Stamp	Single (avg.)	4.5
	Multiple (avg.)	3.25
	Total average	**3.875**
MR Stamp	Single (avg.)	4.25
	Multiple (avg.)	3.75
	Total average	**4**

From these results, we didn't find the effectiveness in the MR stamp system. In the free descriptive answer of the questionnaire, there were several same answers that "It was difficult to find the MR stamp". If the MR stamp can help the broadcaster to find the target object, the broadcaster cannot find the location of the MR stamp itself in the experiment. We expect that an additional function to find the location of the MR stamp improves the effectiveness of the MR stamp system.

6 SECOND EVALUATION WITH SYSTEM IMPROVEMENT

In the first evaluation, there was no difference about the effectiveness between the 2D stamp system and the MR stamp system. Since the prototype system of the MR stamp had a problem that it is difficult to find the location of the MR stamp itself, we try to improve the prototype system to find the MR stamp easily and conduct a second evaluation using the improved prototype system.

6.1 System Improvement

In order to help the broadcaster to find the MR stamp, we introduce spatial audio into the MR stamp. The spatial audio enables the broadcaster to perceive sound all around his/her and be aware of the location of the sound source. Titus [8] studies the effectiveness of spatial audio for finding the location of a maker. He finds that the spatial audio helps the user with a head-mounted display to find a rough location of the maker although it is not enough to specify the exact location of the maker. In our MR stamp system, the broadcaster would find the MR stamp if he/she can be aware of a rough location of the stamp.

We implemented the spatial audio to the MR stamp using a spatial audio function of MRTK (Mixed Reality Toolkit). When the MR stamp is displayed, the spatial sound is generated from the location of the MR stamp. The broadcaster can hear the spatial sound through the HoloLens 2 and find the MR stamp quickly.

6.2 Second Evaluation Results

We conducted a second evaluation using the improved prototype system with spatial audio. The environment and the procedure are same as the first evaluation.

Table 3 shows evaluation result of the second evaluation. The average of time to find the target object in the 2D stamp system takes 17.5 seconds in case of the single object and 19.75 seconds in case of the multiple objects. The total of the average time in the 2D stamp system is 18.625 seconds. In the MR stamp system, it takes 5 seconds in case of the single object and 5 seconds in case of the multiple objects. The total of the average time in the MR stamp system is about 5 seconds. The average of number of errors in the 2D stamp system shows 0.25 and 0.5 in the single and multiple objects respectively. The total of the average number of errors in the 2D stamp system is 0.375. The average of number of errors in the MR stamp system shows 0 and 0 in the single and multiple objects respectively. The total of the average number of errors in the MR stamp system is 0. We conducted a Student's t-test for the results of the total average. There is a significant difference in the average time to find ($p = 0.01$) and a significant trend in the average number of errors ($p = 0.08$) between the 2D and MR stamp systems.

Table 4 shows the result of the questionnaire which is about subjective easiness to find the target object in 5-point scale. The average points of the 2D stamp system are 3 and 2.25 in the single and multiple objects respectively. The total average in the 2D stamp system is 2.625. The average points of the MR stamp system are 4.75 and 4.5 in the single and multiple objects respectively. The total average in the MR stamp system is 4.625. We conducted a Student's t-test for the results of the total average and there is a significant

Table 3: Evaluation result of the second evaluation

Broadcaster	System	Target	Time to find (sec)	Errors (No. of times)
A	2D Stamp	Single	9	0
		Multiple	23	1
	MR Stamp	Single	4	0
		Multiple	5	0
B	2D Stamp	Single	9	0
		Multiple	17	0
	MR Stamp	Single	4	0
		Multiple	5	0
C	2D Stamp	Single	40	1
		Multiple	29	1
	MR Stamp	Single	4	0
		Multiple	5	0
D	2D Stamp	Single	12	0
		Multiple	10	0
	MR Stamp	Single	8	0
		Multiple	5	0
Average	**2D Stamp**	Single	17.5	0.25
		Multiple	19.75	0.5
		Total	**18.625**	**0.375**
	MR Stamp	Single	5	0
		Multiple	5	0
		Total	**5**	**0**

Table 4: Questionnaire result of the second evaluation

System	Target	Subjective easiness to find the target object (5-point scale)
2D Stamp	Single (avg.)	3
	Multiple (avg.)	2.25
	Total average	**2.625**
MR Stamp	Single (avg.)	4.75
	Multiple (avg.)	4.5
	Total average	**4.625**

difference about the subjective easiness to find the target object ($p = 0.0007$).

From these results, we can find the effectiveness in the MR stamp system with special audio. In terms of the three issues of the previous study, the first issue which is that "The stamp cannot be fixed on the target object" is solved because the number of the errors decreases in the proposed system. The second issue which is that "The broadcaster must check a PC display to see the stamp" is solved because the time to find the target object decreases. The third issue which is that "It is difficult to understand a position of the stamp in the real space" is solved because both the number of the errors and the time to find the target object decreases.

7 CONCLUSION

In this study, we proposed MR stamp which can be displayed and fixed on the real space, which enables the broadcaster to check a holographic stamp on the real space through an MR device. We implemented a prototype system of the MR stamp with special audio. In the evaluation of this study, we verified the effectiveness of the MR stamp compared with the 2D stamp using the prototype system. We found that the MR stamp with the spatial audio made it easier for the broadcaster to intuitively grasp the location of the stamp. In addition, a comparison with the 2D stamp system revealed that there was a significant difference in the time to find the target object and the subjective easiness to find the target object. There was also a significant trend in the average number of errors between the 2D and MR stamp systems. From the result, we found that the MR stamp could solve the issues of the 2D stamp. In the future, it is necessary to improve the accuracy of the stamp display.

REFERENCES

[1] Roel Vertegaal: The GAZE groupware system: mediating joint attention in multiparty communication and collaboration, CHI '99 Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pp.294-301 (1999).

[2] David M. Grayson, Andrew F. Monk: Are you looking at me? Eye contact and desktop video conferencing, ACM Transactions on Computer-Human Interaction (TOCHI) Volume 10 Issue 3, September 2003, pp.221-243 (2003).

[3] Yoshia Saito, Aoi Kuzumaki, Yasuhiro Yahata and Dai Nishioka: Development of a Communication Support System with Stamp Functions in 360-degree Internet Live Broadcasting, Proc. of DICOMO2019, pp. 895-900 (2019) (in Japanese).

[4] Gun A. Lee, Theophilus Teo, Seungwon Kim and Mark Billinghurst: Mixed Reality Collaboration through Sharing a Live Panorama, Proc. of SA'17, pp.1-4 (2017).

[5] Janet G Johnson, Danilo Gasques, Tommy Sharkey, Evan Schmitz and Nadir Weibel: Do You Really Need to Know Where "That" Is? Enhancing Support for Referencing in Collaborative Mixed Reality Environments, Proc. of CHI '21, pp. 1-14 (2021).

[6] StateOfVirtualEvents2021, available from <https://s3.amazonaws.com/media.mediapost.com/uploads/StateOfVirtualEvents2021.pdf> (accessed 2022-6-10).

[7] Chris Misa, Dennis Guse, Oliver Hohlfeld, Ramakrishnan Durairajan, Anna Sperotto, Alberto Dainotti and Reza Rejaie: Lessons Learned Organizing the PAM 2020 Virtual Conference, ACM SIGCOMM Computer Communication Review, Vol.50, Issue 3, pp.46-54 (2020).

[8] Titus J. J. Tang and Wai Ho Li: An Assistive EyeWear Prototype that interactively converts 3D Object Locations into Spatial Audio, Proc. of ISWC'14, pp.119-126 (2014).

Object Recognition Method Using Locus of Gestures Detected by YOLOv5

Tsukasa Kudo[†]

[†]Faculty of Informatics, Shizuoka Institute of Science and Technology, Japan
kudo.tsukasa@sist.ac.jp

Abstract - To recognize relatively small objects in an image, it is first necessary to perform object detection. In recent years, research has been actively conducted to utilize deep learning to simultaneously perform object detection and recognition, in which real-time object recognition has been also enabled by such as You Only Look Once (YOLO). However, they are based on deep learning, there are application issues that training data for all targets must be prepared for training the model. In this study, to detect target objects, I propose a method to detect the locus indicated by gestures in videos using YOLOv5, which uses only a single object such as a hand, to extract the target area. In this method, to improve the accuracy of the target area, the false object detection results are eliminated and the locus is corrected, by using the median and moving average of consecutive video frames respectively. Furthermore, it is shown that simple object recognition methods such as template matching can be used by detecting the size and tilt of the target based on the detected area by this method.

Keywords: YOLOv5, Deep learning, Object detection, Gesture recognition, Template matching

1 INTRODUCTION

In recent years, object recognition for videos and images has been actively studied, and its applications are expanding in various fields such as immigration control by face recognition and automatic car driving. However, when the area of the target (hereinafter, target area) in the image is small, it is necessary to perform object detection firstly to specify the target area before object recognition. For example, in the case of face recognition, face detection is performed using Haar-like features, and then face recognition is performed on the detected target area [1].

Using deep learning, various methods have been proposed to efficiently perform both object detection and recognition. For example, You Only Look Once (YOLO) detects the target as a bounding box and simultaneously recognizes the target [2]. Furthermore, it has been shown that even videos can be processed in real-time [3].

However, these methods require the preparation of model training data for each target object, which is a significant burden when the types of target objects are large. On the other hand, for still objects that can be photographed from a specific direction, such as the cover of a book, object recognition can be performed by a simple method such as template matching if the target area can be identified.

In this paper, I propose a method to extract the target area from the gesture in videos by using YOLO for object detection. The gesture is performed so that the target area is a closed area surrounded by the locus of the gesture, and the target area is extracted by using this locus. The important point is that, since gestures can be performed by a certain part of the body such as a hand, only one type of training data is required for a variety of object detections in this method. Also, while other object detection methods extract the target area as a bounding box, this method can extract the target area according to the shape of the target object. That is, for example, when performing template matching, the method can estimate and correct the target's tilt or suppress the influences of background areas.

However, since the above locus is created by continuously detecting a specific part of the gesture in a video, it causes some challenges. The locus contains wrong points due to false detections (hereafter, noises); there may be a double and missing part of the locus at the beginning and end of gestures. To address these challenges, this method eliminates these noises and corrects the locus by utilizing the median and moving average of the locus points. And, I show that this method can extract the target area through experiments.

Furthermore, in order to investigate the effects of size specification and tilt correction on recognition accuracy in template matching, which is one of the simplified object recognition methods, I evaluated the improvement of recognition accuracy for books. The purpose of this evaluation was to clarify the required accuracy in the target area extraction. The results show that when the vicinity of the target region is extracted, with a size error of less than 10% and a tilt error of less than 10°, the detection is correct.

The remainder of this paper is organized as follows. Section 2 presents related works and the aim of this study, and Sec. 3 proposes a target area extraction method based on gestures in a video. Section 4 shows the implementation and experimental results of target area extraction, and Sec. 5 evaluates the accuracy of the proposed method and its effectiveness for template matching. Section 6 discusses on the evaluation results, and Sec. 7 concludes this paper.

2 RELATED WORKS AND AIM OF THIS STUDY

In recent years, the effectiveness of object recognition based on deep learning for images and videos has been widely recognized and applied to various fields. On the other hand, when the target area in an image is relatively small, recognition accuracy deteriorates. So, it is necessary to extract the target area firstly and then perform object recognition.

So, various methods for simultaneously detecting and recognizing objects have been proposed. Faster R-CNN performed both of them in a lump by collective end-to-end train-

ing of both models [4], and YOLO executed them with a single neural network to improve efficiency [2]. Concerning different scale objects, SSD made it possible to process them collectively [5], and RetinaNet improved efficiency by introducing the Feature Pyramid Network (FPN) and improving the loss function [6], [7]. Then, M2Det has further improved accuracy and efficiency by introducing the new FPN and loss function [8].

Among these methods, YOLO has been improved repeatedly through version upgrades, and several models are currently available as YOLOv5 [9]. YOLO estimates the bounding box surrounding the target area and the probability of containing the target when the center of the target area is located in a grid cell. The grid cell is a part of the image divided by grids. And, YOLO is known to have high detection efficiency and accuracy. And, it has been shown that YOLO can be applicable to real-time object detection and recognition [3].

However, because the above methods use deep learning, it is necessary to prepare training data consisting of images and correct labels for model training. For example, YOLO requires not only the preparation of images for training the model but also the corresponding correct labels indicating the location and classification of bounding boxes for each object contained in each image. Therefore, when targeting a large number of object types, the burden of creating these labels increases, which is a major obstacle in practical applications.

On the other hand, trained models and training data for YOLOv5 for various objects, such as the coco dataset, are available on the Internet [9]–[11]. Therefore, when targeting specific objects, YOLO can be easily used for the object detection and recognition from videos in real-time.

The motivation for this study is the idea that the target area of an arbitrary object can be extracted, by detecting the locus of a specific object such as the tip of a hand indicated by a gesture in a video. This locus can be detected in real-time by YOLO, targeting only one type of object, that is, training the model is easy. In addition, using the target area extracted in this way, the target size and tilt can also be estimated from the area. In other words, when recognizing still objects viewed from a specific direction, such as back covers of books on a shelf, it is expected that a simpler method such as template matching can be used instead of the methods using deep learning.

Several applications have been proposed for hand gesture recognition using deep learning, such as conversation and device control [12]–[14]. However, I could not find application studies to extract the target area of an object. Furthermore, in continuous object detection using gestures in videos, it is necessary to eliminate noises due to object detection errors and to correct a double or missing part near the beginning and end of the locus.

The aim of this study is to propose a method for extracting the target area with high accuracy using gesture locus and clarify its effects and practical issues for applying it to object recognition.

Figure 1: Right hand object detection and recognition using YOLOv5

3 PROPOSAL OF TARGET AREA EXTRACTION METHOD USING GESTURE

3.1 Target Video Frame Images

I propose a method for extracting the target area by using gestures in videos. In this method, the locus of the gesture is detected using YOLOv5 (hereinafter, YOLO) sequentially for each frame of the video, and extracts the area surrounded by the locus for the target area. In order to correctly detect the locus indicated by the gesture, we perform noise elimination and locus correction as mentioned in Sec. 2.

Figure 1 shows an object detection and recognition (hereafter, object detection) result image of the right hand by YOLO from a video frame. The detected target is indicated by a bounding box, and the class of the target and the recognition accuracy are indicated above the upper side of the box. In this image, my right hand ("myright") is detected with an accuracy of 77% ("0.77") in the center. On the other hand, the lower-left bounding box is falsely detected noise, and its accuracy is 26%. Therefore, in this case, the former is adopted.

3.2 Target Area Extraction Procedure

The procedure to extract the target area from such as the image of Fig. 1 is shown as follows. Valid coordinates of the locus are selected, and noises are eliminated using the coordinates median of adjacent frames. Then, the locus is corrected using the moving average.

3.2.1 Selection of Valid Coordinates in Locus

In this method, one of the vertices of the bounding box shown in Fig. 1 is selected as a point constructing the locus of the gesture. In the following, it is assumed that the coordinates of the top-left vertex are selected. The valid bounding box is selected in each image with the following condition: its accuracy is the highest in the image and greater than the threshold namely the specified value. And, its coordinate of the top-left vertex is adopted for the locus as the valid coordinate.

If each set of the coordinate and accuracy of the i-th frame is indicated by c_{ij} and a_{ij}, the valid coordinates s_i shown in

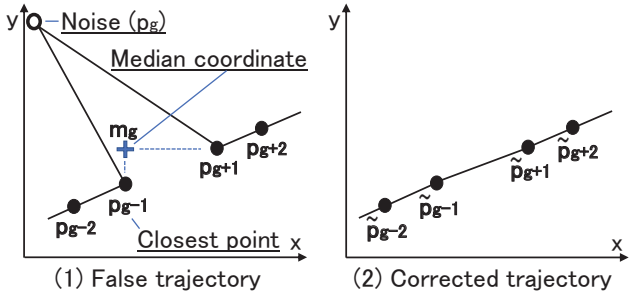


Figure 2: Noise reduction by median coordinate

the Eq. (1) is selected.

$$s_i = \begin{cases} c_{ik} & (\exists k(a_{ik} \geq L \wedge a_{ik} = max(\{a_{ij}\}))) \\ \emptyset & (\forall k(a_{ik} < L)) \end{cases} \quad (1)$$

Here, $\emptyset$ indicates that the coordinates of the frame are not selected; L indicates the threshold; $\{a_{ij}\}$ indicates the set of a_{ij}. In the case of Fig. 1, if $L = 0.4$, then the upper-left coordinate of the center bounding box is selected because its accuracy is the highest and greater than the threshold.

3.2.2 Noise Elimination Using Median Coordinates

To eliminate noises, the median coordinates are calculated from the valid coordinates of the previous and next frames. Let $p_g(g = 1, 2, 3, \cdots)$ be the ordered set of coordinates with eliminating s_i if $s_i = \emptyset$ from $\{s_i\}$ the set of s_i in Eq. (1). Figure 2 (1) shows an example where p_g is a noise.

The median of p_g is constructed using the interval before and after the index g. Let R_g indicates the set of indices of this interval, and let p_{Rx} and p_{Ry} indicate the set of x-coordinates and y-coordinates, respectively. I define the median coordinate of this interval by $m_g = (median(p_{Rx}), median(p_{Rx}))$. Here, *median* is the function to get the median value of the coordinates. And, in the case of Fig. 2, the coordinate "mg" with the median of each of the x-y coordinates is selected.

The noises are eliminated by using these median coordinates. As shown in Eq. (2), if p_g is not the closest coordinate to m_g for the interval R_g, then it is converted to a coordinate $\tilde{p}_g$ with empty $\emptyset$; else p_g is adopted for $\tilde{p}_g$.

$$\tilde{p}_g = \begin{cases} p_g & (\tilde{m}_g = p_g) \\ \emptyset & (\tilde{m}_g \neq p_g) \end{cases} \quad (2)$$

Here,

$$\tilde{m}_g = \{p_h | \exists h(dist(p_h, m_g) = min(dist(p_n, m_g)) \wedge \forall n \in R_g)\}$$

$dist(p_n, m_g)$ indicates the distance between p_n and m_g. In other words, $\tilde{m}_g$ denotes the coordinate of $p_n(r \in R_g)$ that is closest to the median coordinate m_g; and, if the corresponding point p_g is not this coordinate, then it is set to $\emptyset$. The coordinates of the locus without noises are obtained by eliminating $\emptyset$ from the set of coordinates $\{\tilde{p}_g\}$.

In the case of Fig. 2, the closest coordinate to the median $\tilde{m}_g$ is p_{g-1}, so the coordinate $\tilde{p}_g$ is set to $\emptyset$ and eliminated.

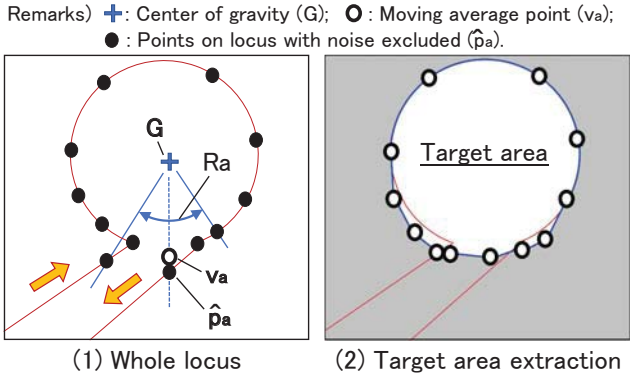


Figure 3: Target area extraction using moving average

The other coordinates of are set as $\tilde{p}_n = p_n(r \in R_g)$. As a result, a locus without the noise is constructed as shown in Fig. 2 (2).

3.2.3 Target Area Extraction by Using Moving Average

To compensate for doubles or missing near the beginning and end of the locus, a moving average of the coordinates along the angle from the center of gravity is created. First, the x-y coordinate of the center of gravity G is obtained as a simple average of the x-coordinate and y-coordinate of the set of coordinates $\{\tilde{p}_g\}$ excluding noises ($\emptyset$), respectively.

Next, as shown in Fig. 3 (1), the coordinates of each point are transformed into a pair $\hat{p}_a = (\theta_a, r_a)$ of angle and distance from G. Let R_a be the interval to calculate the moving average corresponding to $\hat{p}_a$ and define the moving average v_a by Eq. (3).

$$v_a = (\theta_a, \bar{r}_a) \quad (3)$$

Here,

$$\bar{r}_a = (\sum r_u)/n \qquad (u \in R_a)$$

The n is the number of coordinates contained in R_a, and it is 5 in the case of Fig. 3 (1). In other words, $\bar{r}_a$ is the average of the distances between G and the coordinates $\hat{p}_a(a \in R_a)$.

By connecting the coordinates of this moving average set $\{v_a\}$ along the angle, the target area is extracted as shown in Fig. 3 (2).

4 IMPLEMENTATION AND EXPERIMENTS

4.1 Implementation

The experimental system was constructed to verify that the proposed method can extract the target area. This system was implemented on a Windows 10 PC, Python Ver. 3.8.13 as the program, and Pytorch Ver. 1.7.1 with CUDA Ver. 11.5 to use YOLO, OpenCV-Python Ver. 4.5.5.64 for image and video manipulation.

YOLOv5s, a highly efficient model of YOLO, was used and was implemented by adding the necessary functions to the publicly available program [9]. Similarly, the publicly available "Egohand Dataset" [11] was used for the training data of the model. This is the data for training the model to

Figure 4: Gesture locus and target area extraction experiment using proposed method

detect four types of hands, the left and right hands of oneself and the other party, and the number of data is 3,840.

Using a model trained with this data, I implemented a program to extract the target area from a video of hand gestures. First, the hands are detected at each frame, and the valid coordinate in locus is selected using the procedure shown in Sec. 3.2.1. In this implementation, the right hand was used, and the upper-left corner was assumed to be the tip of the hand, as shown in Fig. 1.

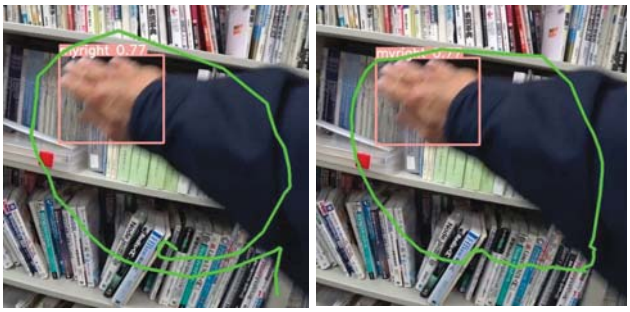
Next, the procedure mentioned in Sec. 3.2.2 and 3.2.3 is used to eliminate noises by median coordinate and extract the target area by moving average. Five points were used to calculate each median coordinate, including the target point and its front and rear points. Since there was no point on one side of the endpoints, their median coordinates are omitted. For the next point, the median coordinate was calculated with three points instead of five points. The moving average is also calculated using 5 points, and the set of moving averages $\{v_a\}$ is obtained. The target area is extracted from $\{v_a\}$ using OpenCV's fillConvexPoly function.

Finally, the bounding box containing the target area is extracted, setting the target area to the frame image and the outside to white.

4.2 Experiments

Using the implemented program, I conducted an experiment to extract the locus of the tip of my hand (hereinafter, hand) captured on video. To confirm that the program can detect even in the case of complex backgrounds, I used the bookshelf shown in Fig. 1. I used a SONY FDRX3000 action camera, and shot videos at $1,920 \times 1,080$ pixels and 30 fps. The hands were moved in a clockwise circular motion starting from the lower right in the image. The accuracy threshold L was set to 0.4.

Figure 4 shows the locus constructed by selecting the coordinates with the highest accuracy for each frame. That is, this is the original locus detected by YOLO. For the background in Fig. 4, the image of Fig. 1 is used. In addition, in the case of this frame image, the hand detected in the center was adopted, so the left-top vertex of its bounding box is on the locus. As shown in Fig. 4, since there were noises due to false detections, the target area could not be extracted directly from this locus.



(1) Locus of median (2) Locus of moving average

Figure 5: Original gesture locus detected by YOLO

Figure 6: Target area extracted by proposed method

Figure 5 (1) shows the locus after selecting only the valid coordinates and eliminating noises by using median as described in Sec.3.2.1 and 3.2.2, respectively. Note that only the vicinity of the locus has been extracted from the whole image. The noises were eliminated, but the locus was doubled near its beginning and end. Figure 5 (2) shows corrected locus from the one shown in Fig. 5 (1) by using the moving average mentioned in Sec. 3.2.3. The moving average corrected the locus to the place between the doubled loci, and a closed area could be constructed. However, near the endpoints, since the hand locus deviated from the target area, an extra area was included as shown right-lower part.

Figure 6 shows the bounding box of the target area extracted using the locus of Fig. 5 (2). The outside of the target area has been transformed to white.

We performed the above procedure three times in the same environment to examine the number of frames detected at each stage of the procedure. Figure 7 shows the results, and

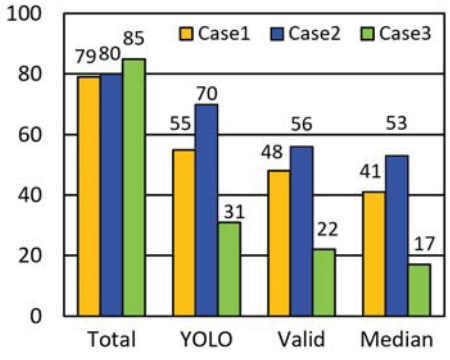


Figure 7: Corrected locus by proposed method

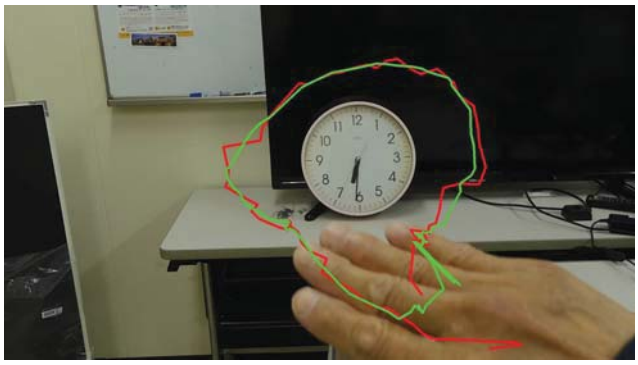


Figure 8: Evaluation of target area extraction accuracy

Figs. 4 to 6 correspond to “Case2”. The vertical axis indicates the number of the detected frames, and the horizontal axis indicates the stage. “Total” shows the total number of frames, “YOLO” shows the number of detected by YOLO including noises, and “Valid” shows the number of accuracies above the threshold (0.4). “Median” shows the number obtained by using medians, that is, the number of frames after eliminating noises, and the number after the moving average is also the same. Note that since the two points at both endpoints of the locus are excluded in the Median stage, as mentioned in Sec.4.1, two points are also excluded for the numbers in the other stages.

As shown in Fig. 7, there was a large difference in the proportion of detections even under similar conditions. In Case3, the proportion detected by YOLO was less than half that of Case2; conversely, Case2 had the highest number of points eliminated due to accuracy under the threshold at the Valid stage. The number of points judged as noise in the Median stage was 7 in Case1 while it was 3 in Case2. The former’s percentage of the total (48), was 14.6%.

5 EVALUATION OF TARGET AREA EXTRACTION AND OBJECT RECOGNITION ACCURACY

5.1 Evaluation of target area extraction accuracy

To evaluate the accuracy of extracting the target area when using hand gestures, I evaluated the extraction accuracy using a round wall clock. In this experiment, the camera was fixed and the hand was moved while watching the monitor in order to evaluate the accuracy of the locus indicated by hand. The used camera was a Nikon COOLPIX A1000, and the resolution and frame rate were the same as in the experiment of Sec. 4.2.

In Fig. 8, the red line shows the locus constructed by using the median; the green line shows the one by using the moving average. The accuracy of the target area is low with respect to the target clock, and in this case, it is outside. Furthermore, at the lower-right part namely near the endpoints of the gesture, it is quite outwardly displaced. The former was caused by using the hand for the gesture, which was too large compared to the target object. The latter was caused by the detection of extra hand movement near the beginning and end of the gesture, namely the movement between the target and the external of the image.

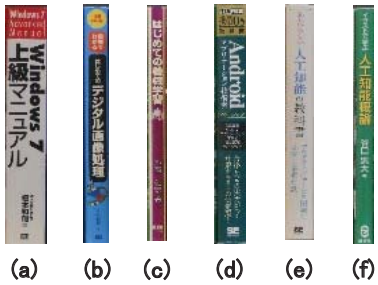


Figure 9: Objects used for template matching

Figure 10: Images for evaluations of size errors

5.2 Evaluation of Object Recognition Accuracy Improvement

To investigate the impact of target area extraction on object recognition accuracy, I evaluated the recognition accuracy of books stored on bookshelves using template matching. The purpose of these evaluations is to clarify the accuracy required to extract the target area by the gesture.

For template matching, the “matchTemplate” function of OpenCV was used with the normalized squared difference matching method. The books to be recognized are the six books shown in Fig. 9. For these books, we evaluated the variation in accuracy when there are errors in size and tilt between the template and the target objects in the image, and when the range of the images was narrowed to the vicinity of the target.

Figure 10 shows the images to evaluate the case of size error. Figure 10 (1) shows the whole image; Fig. 10 (2) shows the image with the narrowed area. Though the latter size is enlarged in this figure, both images are the same size in this experiment. The numbers (a), (b), and (c) below each object correspond to Fig. 9. Similarly, Fig. 11 shows an image to evaluate the case of tilt error, and note that the margins created by the rotation are filled in with white. In addition, the images in Figs. 10 and 11 were shot in a different environment from that of the template image in Fig. 9.

Figure 12 shows the template matching results of all the images in Fig. 9 against Figs. 10 (1) and (2), namely the case of size errors. The “Whole” and “Vicinity” in Fig. 12 corre-

Figure 11: Images for evaluations of tilt errors

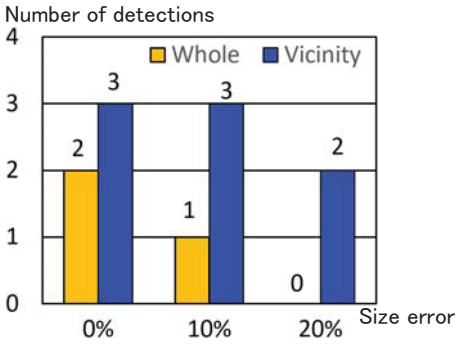


Figure 12: Improvement results for size errors

spond to (1) and (2) in Fig. 10, respectively. “0%” indicates the case where the template size is adjusted to the target of Fig 10, while “10%” and “20%” indicate the case where the template is enlarged to this size, respectively. As shown in Fig. 12, the recognition accuracy degraded as the size error increased, and in the case of the “Whole”, no image was recognized at “20%”. On the other hand, the recognition accuracy in the case of “Vicinity” was improved, and two images were recognized even at “20%”.

Similarly, Fig. 13 shows the result of evaluating the tilt errors, in which “Tilt error” corresponds to the magnitude of the error between the objects in Fig. 9 and Fig. 11. “0°” indicates the case where the image is rotated so that the books are vertical, while “5°” and “10°” indicate the case where the rotation is insufficient by this angle, respectively. As shown in Fig. 13, the recognition accuracy degraded as the tilt error increased, and in the case of the “Whole”, no image was recognized at “10°”. On the other hand, similar to Fig. 12, the recognition accuracy of the “Vicinity” was improved, and two images were recognized even at “10°”.

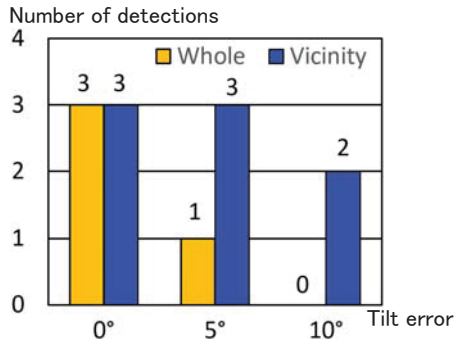


Figure 13: Improvement results for tilt errors

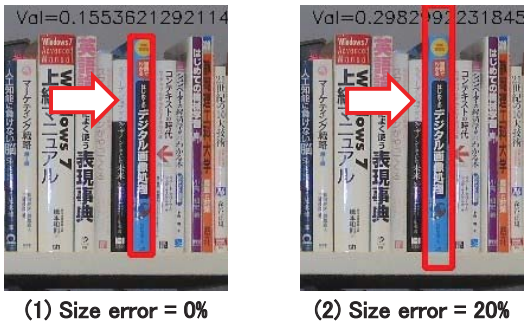


Figure 14: Template matching results on size error

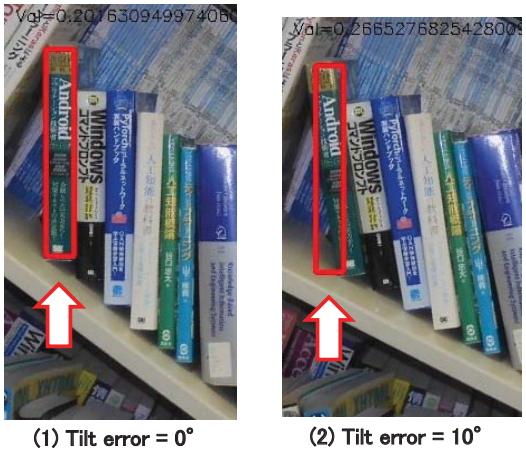


Figure 15: Template matching results on size error

In addition, in the case shown in Fig. 12, the books (d) to (f) shown in Fig. 11 (2) are not detected; conversely, in the case shown in Fig. 13, the books (a) to (c) shown in Fig. 10 (2) were not detected. This is because the tilt of books was too large against the corresponding template shown in Fig. 9, respectively.

Figures 14 and 15 show examples of the matched images for (2) of Figs. 10 and 11, where the matched places are indicated by the rectangles. As shown in (1) of Figs. 14 and 15, when the errors of scale and tilt are small, the extractions were accurate. However, when these errors were large, the accuracies were degraded due to the inclusion of areas other than the target, as shown in (2) of Figs. 14 and 15.

6 DISCUSSION

In this study, I am trying to clarify the effectiveness of the proposed target area extraction method and the issues for its practical applications.

As shown in Fig. 4, several noises tend to be included in the locus due to false object detections used for the gesture. As shown in Fig. 5 (2) and Fig. 8, the proposed method was able to extract the target area by eliminating these noises. Therefore, I consider that the proposed method is effective for

the purpose of target area extraction.

As mentioned in Sec. 2, object recognition accuracy can be improved by specifying the target area in the image, namely by object detection. In addition to this, this method can extract arbitrarily shaped target areas, so it is possible to detect the target size and tilt by utilizing the target area. As shown in Figs. 12 and 13, it was possible to improve the recognition accuracy even in simple object recognition such as template matching by using this data,

On the other hand, through experiments and evaluations, the issues for practical use were found, too. The first issue is object detection accuracy including the position of the object used in the gesture. By extracting the target area using gesture, for example, the "vicinity" shown in Figs. 14 and 15 can be applied for the template matching case. In this case, with a size error of less than 10% and a tilt error of less than 10°, the detection was correct. However, as shown in Fig. 8, the gesture locus deviated from the target area. In addition, as shown in Fig. 7, the object detection accuracy differed greatly for even similar gestures.

The former was due to the hand being too large to specify the target area. The latter was due to the use of existing training data. In other words, it is considered that there were some differences in visibility between the images of existing training data and the images of the gesture, though they are the same objects namely hands.

To address these problems, the following measures can be considered. First, we can use the part of the body that can specify the target area more precisely, such as the fingertip, for the gestures. Second, for efficiently model training, we can use transfer learning based on the model trained with the existing training data used in this study according to the usage part of his/her body. In addition, for distant objects, it is considered effective to perform gestures while monitoring with a wearable camera to suppress the difference in viewpoint between the camera and the operator.

The second issue is that, as shown in Fig. 5 (1), the extra locus is detected. This method aims to automatically extract the target area from the continuously shot videos. However, I found that extra hand motions before and after the target gesture are also detected as a part of the gesture. To address this issue, for example, it is considered to stop the gesture at the beginning and end of the target gesture and identify the extra frames in videos.

The construction and evaluations of these measures are the subjects of the next study.

7 CONCLUSION

In order to recognize a small object in an image, it is first necessary to detect the object, and various methods for simultaneous object detection and recognition such as YOLO have been proposed. However, since these methods utilize deep learning, there is an application problem that it is necessary to prepare the training data for each target object.

For this problem, I propose a method to detect the locus indicated by hand gestures in videos using YOLOv5 and extract the target area. Experiments have shown that this method can eliminate noises due to false detection and improve the accuracy of target area detection. Furthermore, I evaluated the effectiveness of this method for template matching and showed that the recognition accuracy can be improved by detecting the target size and tilt in addition to the target area.

However, it was found that further improvement in the target area detection was necessary to improve this recognition accuracy. So, future studies include improving the accuracy of the target area to be extracted by gestures.

REFERENCES

[1] P. Viola, and M. Jones, "Rapid object detection using a boosted cascade of simple features," Proc. 2001 IEEE Computer Society Conf. Computer Vision and Pattern Recognition, I–I (2001).

[2] J. Redmon, S.Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 779–788 (2016).

[3] Z. Wang, L. Jin, S. Wang, and H. Xu, "Apple stem/calyx real-time recognition using YOLO-v5 algorithm for fruit automatic loading system," Postharvest Biology and Technology, Vol. 185, No. 111808 (2022).

[4] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," Advances in Neural Information Processing Systems, pp. 91–99 (2015).

[5] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," European Conf. Computer Vision, pp. 21–37, Springer. (2016).

[6] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 2117–2125 (2017).

[7] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection." Proc. IEEE Int. Conf. Computer Vision, pp. 2980–2988.(2017).

[8] Q. Zhao, T. Sheng, Y. Wang, Z. Tang, Y. Chen, L. Cai, and H. Ling, "M2det: A single-shot object detector based on multi-level feature pyramid network," Proc. AAAI Conf. Artificial Intelligence, Vol. 33, pp. 9259–9266.(2019).

[9] G. Jpcher, et.al. "YOLOv5," https://github.com/ultralytics/yolov5 (referred May 17, 2022).

[10] T. Y. Lin, et.al., "Microsoft coco: Common objects in context," European Conf. Computer Vision, pp. 740–755 (2014).

[11] S. Bambach, S. Lee, D. Crandall, and C. Yu, "Lending A Hand: Detecting Hands and Recognizing Activities in Complex Egocentric Interactions," IEEE Int. Conf. Computer Vision (ICCV), pp. 1949-1957 (2015), https://public.roboflow.com/object-detection/hands (referred May 19, 2022).

[12] M. Oudah, A. Al-Naji, and J. Chahl, "Hand gesture recognition based on computer vision: a review of techniques," J. Imaging, Vol. 6, No. 8, 73 (2020).

[13] A. Mujahid, M. J. Awan, A. Yasin, M. A. Mohammed,

R. Damaševičius, R. Maskeliūnas, and K. H. Abdulkareem, “Real-time hand gesture recognition based on deep learning YOLOv3 model.” Applied Sciences, Vol. 11, No. 9, 4164 (2021).

[14] Y. Shi, Y. Li, X. Fu, K. Miao, and Q. Miao, “Review of dynamic gesture recognition,” Virtual Reality & Intelligent Hardwar, Vol. 3, No. 3, pp. 183–206 (2021).

Session 4:
Disaster Mitigation
(Chair: Akihiro Hayashi)

An Analysis of Tweets in Disasters for Extracting Rescue Requests

Yuki Koizumi[†], Junji Takemasa[†], Toru Hasegawa[†], and Yoshinobu Kawabe[‡]

[†]Osaka University, Japan
[‡]Aichi Institute of Technology, Japan
{ykoizumi, j-takemasa, t-hasegawa}@ist.osaka-u.ac.jp, kawabe@aitech.ac.jp

Abstract - During catastrophic disasters like the Japan floods 2018, phone-based emergency call systems may not work as expected due to heavy congestion or network disruption. Social media, e.g., Twitter and Facebook, has been playing an essential role as a communication tool to deliver rescue requests in disasters, and it complements phone-based emergency call systems. Understanding rescue requests on social media is key to realizing the automatic extraction of rescue requests from a vast amount of social media posts. This paper analyzes rescue-related tweets, which are tweets containing rescue-related keywords like #rescue, and proposes a taxonomy for rescue-related tweets. According to the analysis, we construct a classifier based on a recurrent neural network and an annotation mechanism to identify why the classifier identifies a tweet as a rescue request. Using these two models, we reveal why there are many false negatives, i.e., the number of rescue requests identified as non-rescue requests.

Keywords: Social Media, Twitter, Disaster, Analysis, Machine Learning

1 Introduction

Delivering rescue requests from citizens in need of help to the right persons, such as rescue authorities and first responders, is a key to effective disaster management. Phone-based emergency call services, however, may not work as expected during and after catastrophic disasters because of network disruption and congestion [1].

Circumstances of rescue requests during disasters have been changing. Social media, e.g., Twitter and Facebook, plays a vital role in delivering rescue requests from victims and it complements existing phone-based emergency call services. Japan, for instance, had several catastrophic floods in 2018 and 2019. One is the *heavy rain of July 2018*, also referred to as *Japan floods 2018* [2] and another is the *19th typhoon of 2019* [3]. We observed that many rescue requests were posted on Twitter during both the disasters.

Few rescue requests on social media, nevertheless, contributed to actual rescue activities. Through our analysis, discussed later in Section 2, reveals that 312 rescue requests were posted during the Japan floods 2018, none of them directly contributed to rescue activities. Most of the rescue requests on Twitter were forwarded to phone-based emergency call services by voluntary workers, which caused further congestion of the phone-based emergency call services. As another example, a local government, Nagano Prefecture, deployed several workers to capture rescue requests from Twitter [4] during the 19th typhoon. On the one hand this activity finally contributed to saving about 50 victims, but on the other hand it consumed many workers, who may have contributed to other tasks. These tales imply that it is indispensable to extract rescue requests on social media automatically to utilize them for rescue activities. Machine learning is a promising technique for filtering rescue requests on social media [5]–[7].

Rescue-related social media posts, which are defined as social media posts having rescue-related keywords like #rescue, are a "mixture of good and bad" as we will analyze them in Section 2. Specifically, many rescue-related social media posts are not related to rescue requests although they have rescue-related keywords. One essential issue for extracting rescue requests from a vast amount of social media posts is variety in contexts of rescue-related social media posts. Rescue-related social media posts include not only rescue requests but also disaster information and sympathy. Another crucial issue is ambiguity in social media posts due to the nature of free-form texts. The ambiguity results in blurring the boundary between rescue requests and disaster information.

Understanding real rescue-related social media posts is a key to realizing good classifiers for extracting rescue requests from social media automatically. While many existing studies analyzed social media posts related to natural disasters, such as hurricanes [8]–[10], few studies analyze rescue requests on social media.

The key contributions of this paper are summarized as follows. First, this is the first study that analyzes rescue-related social media posts from the perspective of extracting rescue requests as far as we know. We captured real social media posts on Twitter during several recent floods in Japan and analyze them. Hereafter, we refer to social media posts simply as tweets since this study focuses on Twitter as social media. Second, we conduct preliminary experiments of classifying rescue requests from numerous tweets with machine learning. Our experiments reveal several lessons to build a classifier based on machine learning to extract rescue requests from tweets.

The paper is organized as follows: In Section 2, we analyze disaster-related and rescue-related tweets. Based on observations found in the analysis, we build a classifier to extract rescue requests from Twitter in Section 3. Section 4 briefly summarizes related work and Section 5 finally concludes this paper.

2 Analysis on Disaster-related Tweets

This section analyzes disaster-related tweets captured during the recent floods in Japan in order to understand rescue requests in tweets.

Table 1: Search keywords used for collecting disaster-related tweets

Category	Keywords
Rain disaster	Heavy rain, rain disaster, disaster, flood, flood disaster, disaster-hit area, flood-hit area, river burst, evacuation, and (confirmation of someone's) safety
Rescue request	Rescue, rescue request, SOS, and help (me)
First responder	Rescue team, fire fighting team, police, Japan self-defense forces, hospital, and local government
Infrastructure	Infrastructure, lifeline, water supply, electricity supply, gas supply, (network) disconnection, (network) congestion, (network) failure, and recovery
Volunteer	Volunteer, support, and relief supply

Table 2: The number of tweets during the Japan floods 2018

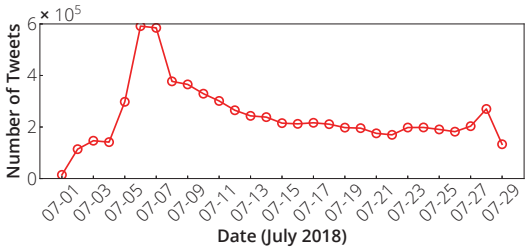
Category	Number
Total	6,978,389
Rescue request	246,807
First responder	932,605
Infrastructure	889,889
Volunteer	324935

2.1 Data Set

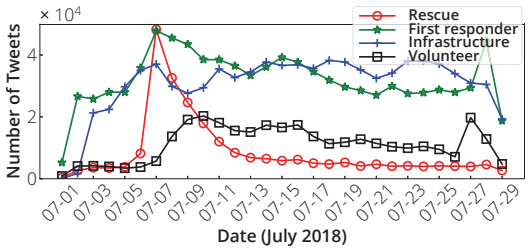
We collected tweets from three flood disasters in Japan: the Japan floods 2018 [2], the 15th [11], and the 19th typhoons [3] of 2019 in Japan. We use the tweets from the Japan floods 2018 for the analyses in this section and those from the other two disasters for the classification in the next section.

The tweets were captured via the Twitter search API [12] by specifying disaster-related keywords, which were selected so that we could capture as many disaster-related tweets as possible. Tweets retweeted with the twitter official API are eliminated from the data set. Let us note that all the keywords are Japanese, and therefore all the tweets are also Japanese. The keywords are categorized into five classes: rain disaster, rescue request, first responder, volunteer, and infrastructure, as summarized in Table 1. We refer to tweets containing at least one keyword in any of the classes as *disaster-related tweets*. In the same way, tweets containing keywords in the classes of rescue request, first responder, infrastructure, and volunteer are referred to as *rescue*, *first responder*, *infrastructure*, and *volunteer-related tweets*, respectively. Since most disaster-related tweets contain keywords in the rain disaster class, we do not focus on tweets in this class. Each class may have tweets unrelated to the class because they were captured with keyword search. For instance, tweets that are not rescue requests, although having rescue-related keywords, are categorized into rescue-related tweets.

The number of collected tweets, i.e., disaster-related tweets, in the Japan floods 2018 is 6,978,389 and the number of tweets in each class is summarized in Table 2.



(a) Number of disaster-related tweets



(b) Number of rescue, first responder, infrastructure, and volunteer-related tweets

Figure 1: Time series analysis of the number of tweets

2.2 Spatiotemporal Analysis

We investigate how the number of tweets grows according to situations in a flood disaster. To understand this analysis, we briefly explain the Japan floods 2018. From the end of June to the middle of July 2018, there was successive heavy rain in western Japan, and it caused widespread and catastrophic floods throughout western Japan, especially in Okayama and Hiroshima. From July 5th to 7th, the most severe floods occurred in Okayama and Hiroshima.

The time series of the number of disaster-related tweets are plotted in Fig. 1a. The horizontal and vertical axes represent the date and the number of tweets in each day. The number of disaster-related tweets steeply increases around July 5th as floods became severe, and its peak is on July 6th and 7th, the most intense days of floods. The numbers of tweets on July 6th and 7th are 591,514 and 584,692, respectively. Note that the number of disaster-related tweets again increases at the end of July because another typhoon came to Japan.

Figure 1b indicates the number of rescue, first responder, infrastructure, and volunteer-related tweets. Though rescue-related tweets also increased as floods became severe, the trend is shifted slightly behind compared to that of disaster-related tweets. That is, the number of rescue-related tweets suddenly increases on July 7th, and the peak is also on the same day. We captured that 48,272 rescue-related tweets on July 7th. First responder and infrastructure-related tweets were captured almost invariably throughout the disaster. In contrast, the number of volunteer-related tweets increases in the aftermath of the disaster.

Next, we analyze the locations where rescue requests were posted. Because few tweets have geographical metadata like geotag information, we extract rescue requests (not rescue-related tweets) that contain a postal address in their text field. In addition, not only retweeted but also copied tweets are eliminated. We found 312 rescue requests indicating a postal address, and we put pins pointing the postal addresses on

Figure 2: A map pointing rescue requests during the Japan floods 2018

a map. Due to privacy issues, we present a coarse-grained map in Fig. 2. One crucial observation is that rescue requests were concentrated in severely flood-hit areas, which were very narrow, in the case of the Japan floods 2018. In a certain town in Okayama, for instance, there were 235 rescue requests indicating a postal address within a 7×5 km rectangle area.

2.3 Understanding Rescue-related Tweets

To understand rescue requests on Twitter, we further picked 5305 tweets, which contains hashtags related with rescue requests, i.e., *#rescue*, *#rescue request*, *#SOS*, *#help*, and *#help_me*, from the aforementioned rescue-related tweets and read all of them. We observe that many of the tweets are not rescue requests. Based on the observation, we define a taxonomy of rescue-related tweets.

Our taxonomy is inspired by the study by Alam et al. [9]. While the taxonomy in [9] categorizes disaster-related tweets, our taxonomy focuses only on rescue-related tweets and analyzes them in detail. Rescue-related tweets are categorized as follows:

- **Rescue request:** A tweet that indicates all the following information: a rescue request, situations of victims, and locations of victims.
- **Incomplete rescue request:** A tweet that indicates a rescue request but lacks either or both the situation and the location of victims.
- **Disaster situation:** A tweet that reports situations of the disaster.
- **Sympathy:** A tweet that prays for the safe rescue of victims.
- **Advice and supplement:** A tweet that gives advice to victims or adds supplementary information to existing tweets.
- **Volunteer:** A tweet that offers or requests voluntary contributions, e.g., voluntary work and donations of supplies like foods, water, and clothes.
- **Exploitation:** A tweet that mentions something unrelated to rescue requests by intentionally exploiting rescue-related keywords so that the tweet is widespread. Such tweets, for instance, include political opinions and advertises services.
- **Unrelated:** A tweet that is not related to the disaster while it contains rescue-related keywords, which are used in a different context from rescue requests, e.g., tweets regarding a smartphone game application.

This taxonomy implies that it is challenging to extract rescue requests via keyword search because many tweets are unrelated to rescue requests while they contain a rescue-related keyword.

According to a recommendation about rescue requests via Twitter posted by Twitter Japan (@TwitterLifeline) [13], victims are recommended to post a rescue request indicating the detailed current situation and the postal address (or geographical location) of victims as well as the hashtag *#rescue*. We, however, captured many rescue requests that lack necessary information like the location of victims. We categorize such tweets into incomplete rescue requests.

We found 312 original rescue requests, some of which do not have necessary information like the location of victims, in the tweet data set. Many voluntary citizens re-posted such an incomplete rescue request by adding missing information, e.g., the hashtag *#rescue*. Hence, there are many tweets categorized to the class of incomplete rescue requests.

3 Classification Based on Machine Learning

In this section, we build a classifier that identifies rescue requests among rescue-related tweets and evaluate the performance. Moreover, we investigate reasons for false negatives with an annotation mechanism [7], which explains reasons why a classifier identifies a tweet as a rescue request.

3.1 Data Set

We use the same data set in Section 2. We use tweets from the Japan floods 2018 and the 19th typhoon for train and test data, respectively, because the two disasters caused more severe damage than recent other typhoons in Japan. In contrast, we use tweets from all the three disasters for creating a corpus of disaster-related tweets because we need as many tweets as possible to create an accurate corpus. For the train and test data sets, we further extract tweets containing rescue-related hashtags, i.e., *#rescue*, *#rescue request*, *#SOS*, *#help*, and *#help me*. The train data set has 2000 rescue-related tweets from the Japan floods 2018 and the test data set has 2091 rescue-related tweets from the 19th typhoon.

3.2 Model of Classifier

We create a classifier using a neural network as a machine learning technique. The classifier consists of three layers: a word embedding layer, a recurrent neural network layer, and a densely connected neural network layer. The word embedding layer converts a word to a vector of real numbers, the recurrent neural network layer learns the relation between tweets and labels, and the densely connected neural network layer summarizes the output of the recurrent neural network layer. A schematic diagram of the classifier is illustrated in Fig. 3.

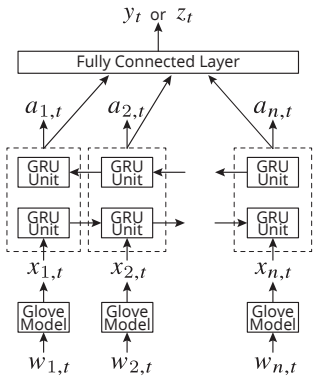


Figure 3: Model of the classifier

3.2.1 Labels of Tweets

A label y_t is assigned to each tweet t according to the taxonomy defined in Section 2. Labels are expressed by 8-bit one-hot binary vectors, where each bit corresponds to each category. For comparison purposes, one-bit binary labels z_t, where 1 indicates t is a rescue request and 0 otherwise. For notation simplicity, the classifier trained with 8 category labels (8-bit labels y_t) and that trained with 2 category labels (1-bit labels z_t) are referred to as the 8-bit and the 1-bit classifier, respectively.

3.2.2 Preliminary Processing of Tweet

A tweet is segmented into words, and the words are converted to their base form with MeCab [14], which is an open-source text segmentation library for the Japanese language. Since spaces do not separate words in the Japanese language, such a tool is necessary for processing Japanese text. Next, URLs, emoticons, and signs, including hash signs, are eliminated. Screen names are converted into a special word, *SCREEN-NAME*. This process converts a tweet t to a sequence of words $\{w_{1,t}, \ldots, w_{n,t}\}$.

3.2.3 Word Embedding

For constructing a word embedding model, we use GloVe [15], which computes vector representations for words according to word-to-word co-occurrence statistics in a corpus. We create a corpus from the tweets from all the three disasters. The corpus has more than 23 million unique tweets, and it contains more than 0.6 billion words. By eliminating words appearing once in the corpus, the corpus finally has 651,423 unique words. We build a 256-dimensional word embedding model from the corpus with GloVe, i.e., a word $w_{i,t}$ is mapped to a vector $x_{i,t} \in \mathbb{R}^{256}$.

3.2.4 Neural Network

We adopt a recurrent neural network as the main body of the classifier. A recurrent neural network is a kind of artificial neural network. It is capable of modeling the dynamics of input values, and hence it is often used as a classifier for time series data, including text data. Kshirsagar et al. [7], for instance, use a recurrent neural network for classifying social media posts. We adopt gated recurrent unit (GRU) cells in our recurrent neural network in the same way as Kshirsagar et al. [7]. As another famous variant of recurrent neural networks, many studies use long short-term memory (LSTM) cells [16]. Both cells can model long-term dependencies among values in an input sequence. For notational simplicity, we refer to a recurrent neural network consisting of GRU and LSTM cells as an LSTM and a GRU network, respectively. Although LSTM networks generally outperform GRU networks [17], LSTM networks have more parameters than GRU networks, and LSTM networks hence require more train data than GRU networks. We compare a GRU and an LSTM network in the following section. Since we do not have sufficient data for training an LSTM network due to the limitation of the number of real rescue requests, we mainly use GRU networks. Finally, we use a bidirectional recurrent network in the same way as the model in [7].

3.2.5 Performance Metrics

We use the following four metrics: accuracy, precision, recall, and F-measure. Let *TP*, *TN*, *FP*, and *FN* be the number of true positives, true negatives, false positives, and false negatives, respectively. In our case, a true positive (negative) corresponds to a (non-)rescue request that is (not) identified as a rescue request. A false negative corresponds to a rescue request that is not identified as a rescue request and a false positive corresponds to a non-rescue request that is identified as a rescue request. The most critical metric for classifiers of rescue requests is recall, which represents the ratio of tweets that are identified as rescue requests among tweets that are actual rescue requests. A low recall value means that many rescue requests will be missed, i.e., there are many false negatives. Since rescue requests must not be missed, recall must be high for classifiers of rescue requests.

3.2.6 Annotation

To understand prediction results made by neural networks, we use an annotation mechanism, also referred to as an attention mechanism [7], [18], which analyzes hidden states $a_{n,t}$ produced by the GRU or the LSTM layer. Hidden states $a_{n,t}$ are referred to as annotation values, hereafter. A high annotation value represents that the corresponding phrase is related to a rescue request.

3.3 Result

We first evaluate the performance of a GRU and an LSTM network by using the 1-bit classifier. Table 3 summarizes the classification results and Table 4 shows the four performance metrics of the two models.

Though both the GRU and the LSTM network achieve high accuracy, they do not achieve high precision and recall. On the one hand the LSTM network outperforms the GRU network in terms of overall performance, i.e., accuracy and F-measure. On the other hand the GRU network is superior to the LSTM network in terms of recall. Since rescue requests must not be

Table 3: Result of classification with the 1-bit classifier

Model	Predict\Label	1	0
GRU	1	209	165
	0	72	1645
LSTM	1	181	70
	0	100	1740

Table 4: Performance of GRU and LSTM

Model	Accuracy	Precision	Recall	F-measure
GRU	0.887	0.559	0.744	0.638
LSTM	0.919	0.721	0.644	0.680

missed, recall must be high. We, hence, use the GRU network in the following experiments. The result, however, shows that the recall of the GRU network is not sufficiently high, and it misses approximately a quarter of rescue requests.

To investigate reasons for the misclassification, we next classify rescue requests into the eight categories, which are discussed in Section 2, using the 8-bit classifier. We use a GRU network in this evaluation. Table 5 summarizes the result. Many rescue requests are misclassified into the following three categories: disaster situation, sympathy, and advice and supplement. There are two typical observations about the misclassification: The first one is that similar expressions are used in both rescue requests and tweets in the three categories and the other one is that many citizens retweeted rescue requests by adding text, which is likely to belong to the three categories.

Regarding the first observation, some rescue requests, for instance, use guessing expressions, which are often used in tweets reporting a disaster situation. The following tweet is a typical example of such rescue requests:

> Help me. My house is flooded above the 2nd floor level. Nobody comes to my rescue. *Roads appear to be blocked due to landslide*.

The last sentence is similar to tweets reporting disaster situations. Such rescue requests are misclassified into the category of disaster situations.

The second observation comes from the fact that there were incomplete rescue requests during the Japan floods 2018, as discussed in the previous section. Many voluntary citizens added several pieces of information, such as hashtags, to incomplete rescue requests to make them complete and retweeted them. On retweeting rescue requests, the citizens often add comments praying for the safety of the victims. The following tweet is a typical example:

> *I hope the victim will be successfully rescued ASAP*. #rescue RT: Help me. My house is flooded.

Another example is as follows:

> *Add #rescue in case you need rescue*. RT: Help me. My house is flooded.

Table 5: Result of classification with the 8-bit classifier

Predict\Label	Rescue request	Non-rescue request
Rescue request	124	28
Incomplete rescue request	83	86
Disaster situation	51	177
Sympathy	11	748
Advice and supplement	8	362
Volunteer	0	1
Exploitation	1	310
Unrelated	3	98

Such tweets are misclassified into the category of either sympathy or advice and supplement.

Finally, we analyze tweets with the annotation mechanism to investigate what kinds of phrases contribute to being rescue requests. We use a 1-bit classifier of a GRU network trained with tweets from the Japan floods 2018. Then, we derive annotation values of phrases, which are generated by deriving all possible n-grams ($3 \leq n \leq 10$) in tweets from both the Japan floods 2018 and the 19th typhoon. The following list summarizes phrases of a high annotation value:

- Characteristics of victims like age and sex (e.g., elderly persons, babies, children, woman, man, octogenarian (80s), septuagenarian (70s), grandfather, and grandmother).
- Situations of flood-hit buildings (e.g., the water level has been rising gradually and my house is flooded above the 2nd floor level).
- Location names of flood-hit areas during the Japan floods 2018 (e.g., Mabi-town, Kurashiki-city, Okayama Prefecture, Japan).
- Numbers in postal addresses (e.g., x-y (x and y represent a block and a house number and they are used like Mabi-town x-y in Japanese style postal addresses)).

In contrast, the annotation values of phrases often used in rescue requests (e.g., rescue request, SOS, help, share it if you can, and retweet it if you can) are very low because they are also used in non-rescue requests. Furthermore, the annotation values of location names of typhoon-hit areas during the 19th typhoon are also very low. If we replace a location name in a false negative rescue request of the 19th typhoon to that of a flood-hit area during the Japan floods 2018, they are identified as a rescue request.

3.4 Lessons Learned

This section summarizes four lessons learned through the classification experiments. First, the entire text of rescue requests should not be used for training classifiers because a part of a rescue request is often unrelated to the rescue request but it is similar to tweets reporting disaster situations. Second, retweeted rescue requests should be carefully handled in the same way as the first lesson because persons retweeting rescue requests often add several texts expressing sympathy for the victims. If texts added on retweeting are unrelated to rescue

requests, they should be eliminated for training a classifier. Third, the entire text of tweets should not use for predicting whether they are rescue requests or not for the same reason behind the first and the second lesson. One way to realize this lesson is to classify tweets using annotation values ($a_{i,t}$ in Fig. 3) as well as a predicted result (z_t) of classifiers. Using annotation values allows us to identify texts related to rescue requests in a tweet and omit texts unrelated to the rescue requests, which cause misclassification, as demonstrated in the previous section. Finally, location names should be handled independently for each disaster. One way to follow this lesson is to convert location names in tweets to a particular reserved word.

4 Related Work

This section compares the present study with related studies.

Several studies analyzed disaster-related social media posts, especially focusing on Twitter. Alam et al. [9] analyzed disaster-related tweets in three hurricanes in the U.S. in 2017, i.e., Harvey, Irma, and Maria. They conducted both textual content analysis and multimedia content analysis on tweets from the three hurricanes. They define a taxonomy for disaster-related tweets and this study inspired us to categorize rescue-related tweets. Yang et al. [10] also analyzed tweets from the hurricane Harvey and proposed a framework to estimate the credibility of events reported by tweets. They also captured disaster-related tweets by searching Twitter for predefined keywords. While those studies focuses on disaster-related tweets, our study focuses on rescue-related tweets.

Next, we introduce studies that proposed classifiers for social media posts with machine learning. Studies in [5] and [6] propose classifiers based on neural networks for detecting fake information or rumors. Though both studies adopt recurrent neural networks, they develop slightly different models. Ma et al. [5] use intervals between social media posts reporting the same event as input values for their classifiers because of the fact that tweets are too short to identify their context with machine learning. Ruchansky et al. [5] propose to use relation between users who post tweets regarding the same event as well as texts of tweets to identify fake news. Kshirsagar et al. [7] propose a classifier for detecting crises like suicide, self-harm, abuse, or eating disorders by using a recurrent neural network. They also develop an annotation mechanism for explaining how social media posts are related to the crises. Their model uses texts of tweets as input values of their classifier. Our model is based on their model.

Finally, we introduce our previous study [19], which develops a communication framework for disaster management. The key idea behind the proposed framework is to utilize social media for collecting information regarding disaster situations. The framework delivers social media posts to right persons by using a machine learning technique. While the motivation of the previous study is to develop a communication framework that utilizes social media, the present study focuses on rescue requests in social media.

5 Conclusion

Circumstances of rescue requests during disasters have been changing. Citizens in need of help use social media, like Twitter, for expressing their rescue requests. To utilize such rescue requests on social media, it is a key to understand real rescue requests on social media. This study captured real disaster-related tweets from several flood disasters in 2018 and 2019 in Japan and analyzed the tweets. We observed that tweets having rescue-related keywords are classified into the eight categories, which include not only rescue requests but also non-rescue requests. Furthermore, most of the rescue-related tweets are unrelated to rescue requests. Next, we conducted preliminary experiments of classifying rescue requests from tweets. We built a classifier based on GRU and LSTM-based recurrent neural networks. The experiments revealed several lessons for building classifiers of rescue requests.

REFERENCES

[1] M. Jahanian *et al.*, "The evolving nature of disaster management in the internet and social media era," in *Proceedings of IEEE LANMAN*, June 2018.

[2] Wikipedia, "2018 Japan floods (heavy rain of July 2018)." https://en.wikipedia.org/wiki/2018_Japan_floods.

[3] Wikipedia, "Typhoon hagibis (the 19th typhoon of 2019 in japan)." https://en.wikipedia.org/wiki/Typhoon_Hagibis_(2019).

[4] Japan Broadcasting Corporation (NHK), "Nagano-prefecture collected rescue requests from Twitter and it contributed to saving about 50 victims." https://bit.ly/3741noG, Nov. 2019. (in Japanese).

[5] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K.-F. Wong, and M. Cha, "Detecting rumors from microblogs with recurrent neural networks," in *Processings of International Joint Conference on Artificial Intelligence*, pp. 3818–3824, July 2016.

[6] N. Ruchansky, S. Seo, and Y. Liu, "CSI: A hybrid deep model for fake news detection," in *Processings of International Conference on Information and Knowledge Management*, pp. 797–806, Nov. 2017.

[7] R. Kshirsagar, R. Morris, and S. R. Bowman, "Detecting and explaining crisis," in *Proceedings of ACM Workshop on Computer Linguuistics and Clinical Psychology*, pp. 66–73, Aug. 2017.

[8] B. Takahashi, E. Tandoc, and C. Carmichael, "Communicating on Twitter during a disaster: An analysis of tweets during typhoon Haiyan in the Philippines," *Computers in Human Behavior*, vol. 50, pp. 392–398, Sept. 2015.

[9] F. Alam, F. Ofli, M. Imran, and M. Aupetit, "A Twitter tale of three hurricanes: Harvey, Irma, and Maria," in *Proceedings of International Conference on Information Systems for Crisis Response and Management*, May 2018.

[10] J. Yang, M. Yu, H. Qin, M. Lu, and C. Yang, "A Twitter data credibility framework—hurricane Harvey as a use

case," *International Journal of Geo-Information*, vol. 8, Feb. 2019.

[11] Wikipedia, "Typhoon faxai (the 15th typhoon of 2019 in japan)." https://en.wikipedia.org/wiki/Typhoon_Faxai_(2019).

[12] Twitter, Inc., "Twitter api documentation—search tweets." https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets.

[13] Twitter, Inc. (@TwitterLifeline), "Twitter post." https://twitter.com/TwitterLifeline/status/1016519147738419201, July 2018.

[14] "MeCab: Yet another part-of-speech and morphological analyzer." https://taku910.github.io/mecab/.

[15] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543, Oct. 2014.

[16] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, pp. 1735–1780, Nov. 1997.

[17] D. Britz, A. Goldie, M.-T. Luong, and Q. Le, "Massive exploration of neural machine translation architectures," *arXiv:1703.03906*, Mar. 2017.

[18] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1480–1489, June 2016.

[19] M. Jahanian, T. Hasegawa, Y. Kawabe, Y. Koizumi, A. Magdy, M. Nishigaki, T. Ohki, and K. K. Ramakrishnan, "Direct: Disaster response coordination with trusted volunteers," in *Proceedings of International Conference on Information and Communication Technologies for Disaster Management*, Dec. 2019.

How can SNS Data Contribute to Disaster Damage Assessment?

Kei Hiroi†, Akihito Kohiga‡, and Yoichi Shinoda‡

†Disaster Prevention Research Institute, Kyoto University, Japan
‡Japan Advanced Institute of Science and Technology, Japan
hiroi@dimsis.dpri.kyoto-u.ac.jp

Abstract - Reports of damage posted on SNS by residents of disaster-striken area at the time of a disaster are expected to be of a great use. They may be a valuable source of information in areas where it is difficult to install, operate, and maintain observation device, or where devices are just missing. However, it has not yet been determined how they can be used effectively for damage assessment. Therefore, a study on a complementary use of SNS data for flood analysis using data assimilation to improve damage assessment is in urgent demand. In this paper, we report the evaluation results of data assimilation assuming that SNS data can be collected stably, and discuss about the impact of error values.

Keywords: flood estimation, state-space model, temporal-spatial analysis, data assimilation

1 Introduction

There are concerns that the risk of floods on a global scale will intensify. The IPCC's Fifth Assessment Report stated that global warming is gradually progressing, and it is likely that the frequency and intensity of rainfall will change accordingly [1]. There are already many areas where the frequency and intensity of heavy rain and flooding are increasing worldwide [2] [3]. As part of measures against flood damage, observing rainfall, river, and flooding, and grasping the changing situation of rainfall and river and its influence enables people to judge what action they should take and take effective steps to prevent or mitigate damage. Many previous studies have attempted to estimate flood risks using the vulnerability of the area. For example, in [4], the flood risk in the city was estimated with a detailed spatial resolution of approximately 2 meters. [5] [6] conducted research to estimate index-based flood risk using a theoretical hydraulic engineering model. Furthermore, a Chinese case in [7] examined recognizing risks with the situation of 1997's Red River flood. Studies are actively conducted to correctly analyze risks by presenting risks to people in affected areas and raising awareness of individual flood risks to lead to mitigation behavior [8] [9].

However, these previous studies are not temporal estimation methods but rather only static estimation approaches to calculate the maximum water level. Such a static estimation result can be said to be a risk estimation in which the risk value might change due to fluctuations in rainfall. Considering evacuation behavior, dynamic risk estimation is required because flood situations change very rapidly with the flooding phenomenon over streets due to the water overflowing from small rivers and waterways spreading throughout the city in a complicated manner and due to the water from the rain that cannot be completely drained. Therefore, it is necessary to calculate the high temporal-spatial flood level estimation that fluctuates according to the rainfall situation to know the risk with a high temporal resolution for guiding evacuation behavior. Our research goal is to detect flooding as time-series data with only a limited number of observation devices.

This paper investigate whether SNS data can be used to assess flooding. SNS data is said to be effective to figure out flood levels even in places where it is difficult to install, operate, and manage observation devices. Although there have been many studies on flood damage detection using SNS, their effectiveness has not been clarified, and the amount and content of data collected are not fixed depending on the case of flood damage. This paper validates the SNS data using the following procedure based on our state-space model(SSM), which is our previous research. The purpose of SNS Data Validation is to investigate whether SNS data can contribute to the accuracy of flood assessment and under what conditions SNS data can improve the accuracy.

Then we also suggest a system for an appropriate SNS use case that could improve the accuracy of flooding assessment further by adding SNS data as well as observation data, and calculate flooding conditions for the entire affected area. In order to realize this system, it necessary to validate the effectiveness of the SNS data.

We generate quantified SNS data from flood simulations in this paper. To generate the SNS data, we use time-series data collected from observation devices at multiple locations. The results of flood analysis simulation were assimilated using the time-series data to simulate the SNS locations, the timing of postings, and numerical flood levels . Then, we append errors to the simulated SNS data. Afterward, we regenerate a flood level on the observation location based on the simulated SNS data with errors and examine the accuracy of the errors on the data assimilation accuracy.

2 Related Works

2.1 Flood Monitoring

Traditional river sensors [10], [11] have succeeded in detecting disaster signs of large-scale rivers, which have both the advantage of stable monitoring and the disadvantage of installa-

tion limitation (i.e., very big equipment, high installation cost of several million dollars, and complicated pre-configuration). Improvements in installation limitations suggest the possibility of a large number of sensor installations and reliable detection to improve monitoring sensors with higher resolution. Approaching flood prediction, hydrological techniques [12] or Artificial Neural Networks [13], [14] are proposed as high prediction methods. Predicting rising river levels has resulted in highly precise river inflow in the view of large-scale river analysis. However, these previous methods cannot predict the flooding of smaller rivers and waterways. This is because complex water flow prediction requires analyzing complicated relationships among a plurality of confluent rivers and factoring in the impact of rainfall dynamics.

2.2 Risk Estimation with Higher Spatial Resolution

Various studies have already attempted to generate information on which place is dangerous. In case studies such as [15] [16], research aimed at presenting the risk on a map. Sinnakaudan et al. [15] developed an ArcView GIS extension as an efficient and interactive spatial decision support tool for flood risk analysis. Their extension has the capability to analyze the computed water surface profiles and produce a related flood map for the Pari River in ArcView GIS. As another GIS-based flood risk assessment, Lyu et al. [16] studied the Guangzhou metro system's vulnerability. Their results showed the vulnerability of several metro stations using the flood event that occurred in Guangzhou on May 10, 2016.

Some studies have already proposed modeling methods that strictly collect data as input data [4] [17] [18]. Ernst et al. [4] presented a microscale flood risk analysis procedure as a 2-meter grid, relying on detailed 2D inundation modeling and on a high-resolution topographic and land-use database. However, detailed risk estimation requires detailed data measurement, such as laser altimeter data, and it is not realistic to measure such data in all areas.

2.3 Flood Ditection through Social Networking Service

One other way to learn about flooded areas is through social networking services (SNS). Kim et al. [19] stated that social networking is the fourth most popular information source for accessing emergency information. Then they applied social network analysis to convert emergency social network data into knowledge for the 2016 flood in Louisiana. Their objective was to support emergency agencies develop their social media operation strategies for a disaster mitigation plan. This reference explores patterns of interaction between online users and disaster responses.

Sufi et al. [20] designed a disaster monitoring system on social media feeds related to disasters through AI- and NLP-based sentiment analysis. Their system has a mean accuracy of 0.05. They report that their system shows potential disaster locations with an average accuracy of 0.93. Teodorescu [21] designed a method for the analysis of SNS for the purpose of forecasting needs and measures for relief and mitigation. His method analyzes SNS related time-series with the aim of establishing correlations between characteristics of the disaster and the SNS response.

2.4 Issues and Approaches

In order to comprehend safe evacuation routes, it is important to figure out the situation with regard to roads in urban areas. Currently, flood damage assessment is based on two methods: numerical simulation (e.g., flood analysis) and monitoring using low-resolution ground observation data (precipitation and river water levels).

Numerical simulations are based on differential equations for flood flow in urban areas for a given amount of precipitation, and the maximum flood level in a detailed space (e.g., a 10-m grid) is calculated. Based on the calculated results, hazardous areas in the event of heavy rainfall are published. However, the analysis uses an ideal model that assumes fixed parameters such as the amount of precipitation, its runoff coefficient, and the outflow conditions of drainage channels. Therefore, in urban areas with complex rainfall distribution and land use, the analysis results and actual flood levels will differ. As the result, flooding of roads occurs prior to the announcement of warnings and evacuation information, leading to damage.

On the other hand, monitoring establishes thresholds for dangerous water levels at specific locations where there is concern about road underpasses and river breaches. This is a situation monitoring method to detect the occurrence of flooding based on observation data. This method is easy to assess the actual damage but has the problem of limited observation points.

SNS data is expected to solve these monitoring limitations. As indicated in the previous section, the importance of SNS in flood damage detection has long been known and has been applied in many flood damage cases. However, there is a fundamental problem with water damage detection using SNS. That is, SNS data is not necessarily posted in every case. While it may work effectively in floods with a high number of postings, it is highly likely that it will not be as accurate as reported in floods with a low number of postings. In particular, it may be difficult to post while ensuring safety in heavily damaged areas, and communication problems may prevent posting. We convinced that these problems are obstacles to the effective use of SNS for flood damage detection. Therefore, in this paper, we investigate how much SNS data on the number of postings, their contents, and the timing of postings would be effective for damage assessment.

We intend to develop a system for improving the accuracy of flood level estimation through data assimilation using heterogeneous data as a prospect for the investigation in this paper. We have previously proposed a method for estimating the expansion process of flooding by applying data assimilation using

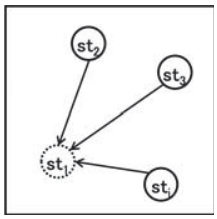
observation time-series data as heterogeneous data to simulate flood analysis. Our estimation method showed a significant improvement, with an error of less than 9 cm. We are planning to add SNS data to this estimation method to further improve its accuracy and to develop a system to present the flood disaster situation of the entire affected area. However, although there have been many studies on flood damage detection using SNS, the amount and content of the data collected are not well defined for each flood damage case. The effectiveness of using SNS for flood level estimation is not yet clarified. Therefore, with the aim of developing the system to accurately estimate and present the flooding situation of the entire affected area, this paper investigates the relationship between SNS data and the accuracy of flooding estimation through simulations.

3 Proposal of Data Assimilation Method for Improving Accuracy of Flood Estimation

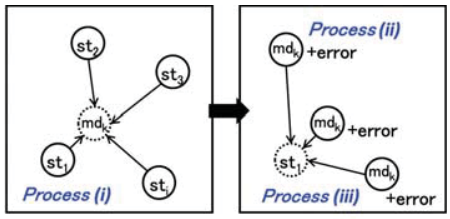
3.1 SNS Data Validation for Data Assimilation

The basic idea of SNS Data Validation for Data Assimilation in this paper is illustrated in Figure 1(a), state-space model: SSM using SNS Data. The purpose of SNS Data Validation is to investigate whether SNS data can contribute to precise estimation of flood assessment and under what conditions SNS data can improve the accuracy. Since we are unable to determine the error rate contained in the SNS data collected at the time of flooding, we generate altinative quantified SNS data from simulations. To generate the SNS data, we use time-series data collected from observation devices at multiple locations. The results of flood analysis simulation are assimilated with the time-series data to simulate the SNS locations, the timing of postings, and numerical flood levels (Figure 1(b)(i)). For this data assimilation, we used a spatial-temporal state-space model proposed in a previous our study [22](Figure 1(a) SSM using Observation Data). Note that, a spatial-temporal state-space model in this paper, is applied without using the waterway and sewer data used in the previous study [22], due to these data are generally limited in availability.

Then, we append errors to the simulated SNS data (Figure 1(b)(ii)). This assumes the errors in location, time, and water level value that are present in the textual data of the actual SNS data. Afterward, we regenerate a flood level on the observation location based on the simulated SNS data with errors and examine the accuracy of the errors on the data assimilation accuracy (Figure 1(b)(iii)). This regeneration uses a state-space model that applies the state-space model of the previous study [22] toward the spatial direction. Here, if there is a small difference between the time-series data and the data assimilation results on observation location, it can be expected that the SNS data is applicable to flood assessment by data assimilation. Contrary to this, if the data assimilation accuracy is low despite the small error appended to the simulated SNS data, then there are problems in the use of SNS data.



(a) SSM using Observation Data



(b) SSM using SNS Data

Figure 1: SNS Data Validation for Data Assimilation ($st_{1,\ldots,i}$:observation location, md_k:SNS data posting location)

3.1.1 Process(i) Simulated SNS data for Flood Analysis Simulation

The process flow is shown in Figure 2. In process(i), the results of the flood analysis simulation are assimilated with time-series data collected from observation locations to simulate the locations, timing of posting, and flood water level values. This section outlines the spatial-temporal state-space model used in our data assimilation. The basic flood analysis is based on the conventional simulation using a surface flooding model. This method calculates the amount of runoff at each grid location by expressing the flooding flow as a continuous equation and motion equations.

A continuous equation is defined as follows.

$$\frac{\partial h}{\partial t} + \frac{\partial M}{\partial x} + \frac{\partial N}{\partial y} = 0 \tag{1}$$

Motion equations are

$$\frac{\partial M}{\partial t} + \frac{\partial UM}{\partial x} + \frac{\partial VM}{\partial y} + gh\frac{\partial H}{\partial x} + \frac{1}{\rho}\tau_x(b) = 0 \tag{2}$$

$$\frac{\partial N}{\partial t} + \frac{\partial UN}{\partial x} + \frac{\partial VN}{\partial y} + gh\frac{\partial H}{\partial y} + \frac{1}{\rho}\tau_y(b) = 0 \tag{3}$$

Each parameter is defined as t: time, H: water level, h: flood level, U: flow velocity (X direction), V: flow velocity (Y direction), g: gravity acceleration, ρ: water density, M: flux (X direction), N: flux (Y direction) ($M = uh$, $N = vh$).

Here, shear force in x direction $\tau_x(b)$ and shear force in y direction $\tau_y(b)$ are defined as follows.

$$\tau_x(b) = \frac{\rho g n^2 \overline{U}\sqrt{U^2 + V^2}}{h^{\frac{1}{3}}} \tag{4}$$

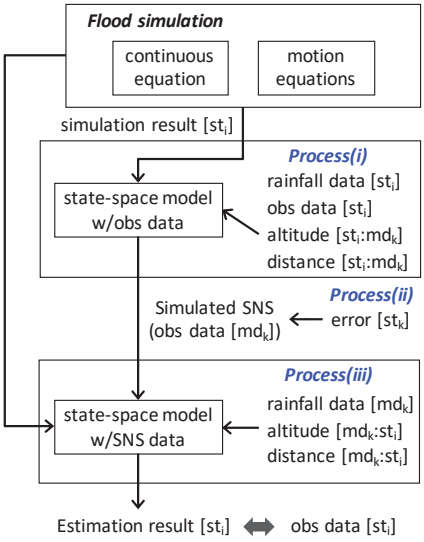


Figure 2: SNS Data Process Flow ($st_{1,\ldots,i}$:observation location, md_k:SNS data posting location)

$$\tau_y(b) = \frac{\rho g n^2 \overline{V}\sqrt{U^2+V^2}}{h^{\frac{1}{3}}} \tag{5}$$

The roughness coefficient n (The resistance value of river water to touch obstacles) can be expressed as follows, taking into account the influence of the building.

$$n^2 = n_0^2 + 0.020 \times \frac{\theta}{100-\theta} \times h^{\frac{4}{3}} \tag{6}$$

(n:bottom roughness coefficient, n_o:composition equivalent roughness coefficient,θ:building occupancy rate)

Equation (1)-(3) calculates flood level h for each grid, accounting for the runoff from the inside of the sewer line to the ground surface and the flooding to the ground surface from rainwater. For the above equations, the inflow into each grid represents the flux into each grid from adjacent grids and the effect of buildings on the inflow in each grid.

We define D as the two-dimensional space corresponding to the region of interest and divide D into m grids of d meters each. Let $s_i \in D$ denote the location coordinates of each grid (s_i is denoted by i). Using equation (1)-(3), we calculate $h_t(i)$ for flood level of each grid at time t.

Then, using the state-space model, we estimate the flood level of grid s_k from the observations $y_t^{(i)}$ collected at the observation location at time t. Grid $s_k(k = 1, 2, 3...., m)$ is the location indicated by the SNS data. The state-space model is represented by two types of observation equations; the flood analysis simulation result $h_t(i)$ at grid s_i, and the difference between the flood analysis simulation and the observed value at the observation location. This state-space model is defined by the equations (7)(8)(9).

$$\boldsymbol{y_t^{(i)}} = S_t \boldsymbol{r_t^{(i)}} + \boldsymbol{G_t^{(i)}} \boldsymbol{x_t^{(i)}} + \boldsymbol{e_t^{(i)}} \tag{7}$$

$$\boldsymbol{r_t^{(i)}} = \boldsymbol{r_{t-1}^{(i)}} + \boldsymbol{v_t^{(i)}} \tag{8}$$

$$\boldsymbol{x_t^{(i)}} = \boldsymbol{x_{t-1}^{(i)}} + \boldsymbol{u_t^{(i)}} \tag{9}$$

The $\boldsymbol{r_t^{(i)}}$ denotes the state at time t and $\boldsymbol{v_t^{(i)}}$ denotes noise. The term $\boldsymbol{G_t^{(i)}} \boldsymbol{x_t^{(i)}}$ represents the total inflow/outflow, and $\boldsymbol{x_t^{(i)}}$ is the difference between the flood analysis simulation results and the observed values. The $\boldsymbol{u_t^{(i)}}$ denotes the noise at time t. $\boldsymbol{G_t^{(i)}}$ is the adjacency matrix indicating the spatial component.

3.1.2 Process(ii) Appending Errors to Simulated SNS Data

Now, in process(ii), we append the error component to the simulated SNS data. Actual SNS data shows a variety of representations of water levels. For example, “It’s flooded up to my knees," “The car is flooded," or pictures are posted with comments such as “It’s raining so hard." This paper considers SNS data that expresses water levels in words. The flood level $h_t(k)$ indicated by the SNS data is assumed to contain an error component e. The representation type of the SNS data is considered as a factor that causes errors due to the quantification of the SNS data (Table 1). We assume that the type of data representation occurs for each of the posted location, time, and water level values. For water level values, SNS data can be expressed in the form of measurement, comparison with an object, or description of the situation. In the case of measurement, it is considered to be measured at a guess, which results in a difference from the actual water level.

When comparing with objects, water levels are explained based on objects such as knee height or up to the ankles, but the sizes of these objects vary from each user, so even if quantified, they differ from the actual water level. Although this is only an assumption, the average length below the knee for Japanese people is 46.7 cm in males and 42.9 cm in females, a difference of about 4 cm even in the average value. In the description of the situation, the data explains mostly describes the aspect of flooding, with little mention of water levels, in consequence, it is expected that quantification itself is often difficult.

For information representation of location, the following information can be considered: GPS, address, road/river, landmark, and city/town name. When GPS is given to SNS data, the exact location at the time of posting can be reflected in the quantified SNS data. However, if the location indicated by the posted message differs from the location at the time of posting, there is an error from the posted location. The same error occurs for other types of data representations. In some cases, addresses of flooded areas are posted for rescue in flooding situations. Although there may be an error of a few meters, the location information would be approximately correct. If a road/river is described as a location, it is considered difficult to determine the exact location from the text content itself. In the case of landmarks, the data may indicate the location in front of the landmark, whereas it is also possible that the data indicates the

Table 1: SNS Data Representation Types

Data	Types	Example
Level	measurement	30cm
	comparison	knee height
	description	looks like river
Location	GPS	34.9104, 135.8002
	address	1-1 Gokasho,Uji-city,
	road/river	Route 24
	landmark	In fromt of Kyoto Station
	city/town name	Uji city
Time	timestamp	2022/7/8/21:00
	comparison	just now
	range	about 21:00

location where the landmark is visible, in which case a large error of about 100 meters or more would occur. For city/town names, we consider a significant error of several kilometers in identifying the location, due to the wide range of areas indicated by the data.

For the representation of time information, there are three forms to consider timestamp, comparison, and date/time range. The timestamp can show the exact date and time in the simulated SNS data. As for comparison, it is considered to be a popular form of time, nevertheless representations such as "just now" are likely to include an error of several tens of minutes, as the sense of time differs among individuals. Additionally, it is assumed that there are many cases describing a range of dates/times, such as "around 21:00" or "this evening". While errors can be expected to be small for numerical time representation, in the case of "night" and other representations, errors are likely to be in the order of several hours. Furthermore, as with location, there are cases in which the time of posting also differs from the time of flooding, and this difference may result in a significant error in the time representation.

We define the error components at location s_k, resulting from the quantification, as the error in the representation type $e_{\Delta v,\Delta l,\Delta t}(k)$, the error relative to the posting location/time $\zeta_{\Delta l'}(k)$, and the error from the time of posting $\zeta_{\Delta t'}(k)$.

3.1.3 Process(iii) Flood Level Estimation and its Validation

Process(iii) regenerates the time-series data for the observation location to investigate the error impact on the data assimilation accuracy based on the simulated SNS data with error $h_t(k) + e_{\Delta v,\Delta l,\Delta t}(k) + \zeta_{\Delta l',\Delta t'}(k)$. Here, we use a state-space model that utilizes the model detailed in 3.1.1 in the spatial direction. For equation (7)(8)(9), at time t, the simulated SNS data $h_t(k) + e_{\Delta v,\Delta l,\Delta t}(k) + \zeta_{\Delta l',\Delta t'}(k)$ is substituted into $y_t^{(i)}$ to estimate the flood level $h'(k')$ of the target location $s_{k'}$. Simulated SNS data on a particular location is not continuous time-series data. Thus, there is only one t in the simulated SNS data, and the state-space model in process(iii) is applied only in the spatial direction.

The difference between the restored water level $h'(k')$ and the actual water level $h(k')$ is shown as the effect of the error component $e_{\Delta v,\Delta l,\Delta t}(k) + \zeta_{\Delta l',\Delta t'}(k)$ on the data assimilation method. This paper validates the SNS data effectiveness by comparing flood level $h'(i)$ regenerated from the simulated SNS data at location k with the actual observed water level $h(i)$.

3.2 Evaluation and Discussion

3.2.1 Evaluation Purpose and Flood Case

This evaluation investigates the effect of the data representation that text-based SNSs have on the quantification of flood levels. SNS data reporting water levels during floods often include measurement of water levels or are expressed in comparison to a body of water. In addition to water levels, location and time also contain errors in quantification, depending on the type of representation. We investigate the impact of SNS data with these errors on statistical flood analysis, such as data assimilation. Based on the results, we will discuss what form of SNS data would contribute to flood assessments.

Our evaluation is based on SNS data simulated by a state-space model with observed data of flood. We append errors to the simulated SNS data and observe the effects of the errors. Then, we determine errors that could occur due to the quantification of the SNS data, as shown in Figure 3. This paper uses flood observations collected in Tsushima City, Aichi Prefecture, Japan, from October 22 to 23, 2017, due to rainfall caused by Typhoon No. 23. A rainfall of 32 mm/h was observed at 23:00 at the nearest precipitation gauge (Aisai Observatory, Aichi Prefecture) to our target area. Using this rainfall data as input values, we calculated a flood analysis simulation with the flood analysis simulation NILIM 2.0. With the results of the simulated flood analysis, we apply the flood estimation method using the state-space model. For generating simulated SNS data, the state-spatial model with observation data uses water level observation data at waterways every 5 minutes collected from pressure-type sensors installed at four locations in the target area.

Detailed information about the observation locations is presented as follows. Observation locations 1 and 2, and 3 and 4 are on the same waterway. The distance between observation locations 1 and 2 is approximately 500 meters, and the distance between locations 3 and 4 is approximately 600 meters. Observation location 4 is connected to the sewer. Two waterways are approximately 500 meters apart. There are no floodgates between the observation locations. The difference in elevation in this area is within 0.30 meters, and the elevation values (Elevation model by Geospatial Information Authority of Japan) are equal at all four observation locations. The heights from the bottom of the waterway to the road are 1.01, 1.14, 0.72, and 1.28 meters, and the usual water levels are 0.06, 0.07, 0.16, and 0.31 meters. On the day of the flood, the installed sensor devices showed, water overflowing the waterways and flood levels of up to 0.26, 0.25, 0.63, and 0.48 meters above the road.

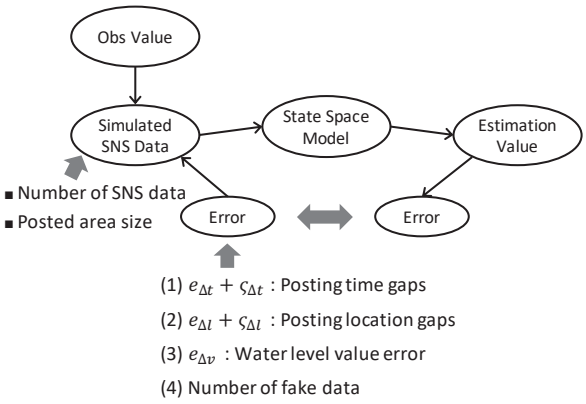


Figure 3: Appending Errors to Simulated SNS Data

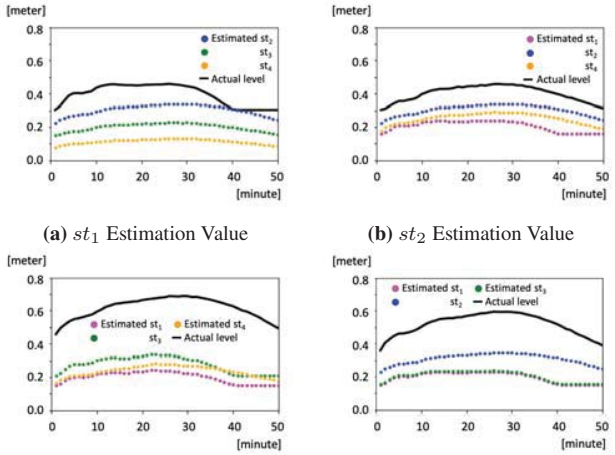
3.2.2 Evaluation Procedure

The evaluation randomly extracts simulated SNS data according to the Number of SNS data (8,24,48,80,120,435) from the Posted area size (10,20,30,40,50, 100[meters]). Errors appending to the simulated SNS data are parameterized to indicate fluctuations. The parameters that indicate the fluctuation of each error are 0,10,20,30[cm]. We begin by using the number of SNS data posted and area size where SNS data were posted as two common parameters. Subsequently, we adopt the parameters to generate the errors as water level errors. The error, resulting from the information representation of the water level value, could be the fluctuation of the value when the water level is estimated from the posted SNS data. Although this fluctuation can have a range of different values, this paper assumes a fixed parameter for the error in order to verify the effect of the SNS data. We add the value indicated by the parameter: water level value error to the water level in the simulated SNS data.

We repeat the above procedure five times and compute the mean value of the estimation. Simulated SNSs appended with the errors are applied to the state-space model to calculate the estimated water levels for the four observation sites with time-series data.

Preliminary result: SSM with Observation Data For comparing the accuracy, this section shows the estimated flood level $l^{(i)}_{tT,k}$ using the state-space model with observation data (Figure 1 SSM using Observation Data). Of the four observation locations, we apply the observation data from three locations (from time $t = 0$ to $t = 50$) to "SSM using Observation Data" to estimate the flood water level at the remaining one location (the water level at st_1 is estimated using the water level time-series data at st_2, st_3, and st_4). The estimation results are shown in Figures 8. The black lines indicate the actual observed flood water levels, and the magenta, green, blue, and orange dots indicate the estimated water levels using the state-space model. Root Mean Squared Error: RMSE between the mean of the estimation results and the actual observed values is shown in Table 2.

In Figure 8 (a), the estimated water level at st_1 was the most accurate using the data from st_3, with an average RMSE of 0.19 meters. The estimated water level for st_2 in Figure 8 (b) was also the most accurate, with an average RMSE of 0.20 meters, using data from st_3. Both estimation for st_3 in Figure 8 (c) and st_4 in and Figure 8 (d) using data from other observation locations were less accurate, with average RMSEs of 0.38 meters and 0.28 meters. The results of either estimation resulted in a large difference from the water level at the flood peak. These results are due to the failure of the "SSM using Observation Data" to follow the rising and falling water levels. In our previous study [22], the maximum error was 9 cm because we included data of waterways and sewers. However, this paper does not use those data, in order to apply our system in areas where it is difficult to collect the data. For all observation locations, we found that estimation using data from distant observation locations resulted in large errors.



(a) st_1 Estimation Value **(b)** st_2 Estimation Value

(c) st_3 Estimation Value **(d)** st_4 Estimation Value

Figure 4: Preliminary result: SSM with Observation Data

Table 2: RSME: SSM with Observation Data

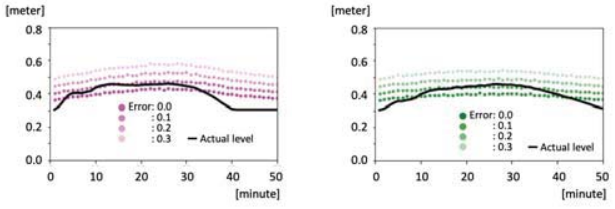
Location	mean	minimum	maximun
st_1	0.19	0.12	0.24
st_2	0.20	0.15	0.23
st_3	0.38	0.28	0.43
st_4	0.28	0.18	0.33

Result This section shows the results of applying the state-space model with error values appended to the simulated SNS data as Water Level Value Error in Table 3. When no error values were added, the mean values of st_1 and st_2 showed the smallest RSME. The larger error value resulted in a larger RSME. On the other hand, st_3 and st_4 showed a large RSME in mean value.

Figure 13 compares the time-series data of the actual observations with the mean value of the estimated values. For st_1 and st_2, as the water level changes, the water level estimated from the SNS data also changes and shows little difference from

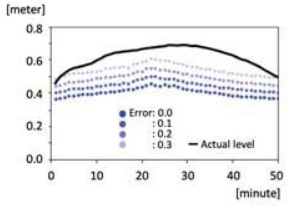
Table 3: RSME: Result(3)：Water Level Value Error

Value Error [meter]		0.00	0.10	0.20	0.30
st_1	mean	0.04	0.05	0.10	0.15
	minimun	0.00	0.00	0.01	0.06
	maximun	0.13	0.17	0.27	0.30
st_2	mean	0.04	0.03	0.07	0.12
	minimun	0.00	0.00	0.00	0.03
	maximun	0.26	0.15	0.25	0.39
st_3	mean	0.21	0.17	0.13	0.08
	minimun	0.00	0.03	0.00	0.00
	maximun	0.30	0.28	0.25	0.22
st_4	mean	0.13	0.09	0.06	0.05
	minimun	0.00	0.00	0.00	0.00
	maximun	0.22	0.18	0.37	0.52

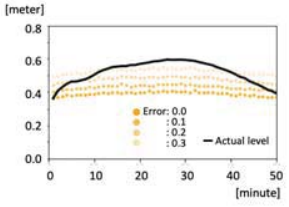


(a) st_1 Estimation Value

(b) st_2 Estimation Value



(c) st_3 Estimation Value



(d) st_4 Estimation Value

Figure 5: Result(3)：Water Level Value Error (Mean Value)

the actual observation. In addition, larger error values tend to provide larger estimates of the results. Meanwhile, ST_3 and ST_4 result in a difference of about 0.20 meters between the estimated result (error value: 0) and the actual water level at the peak of the flooding. We consider that the estimated values were closer to the actual observed value when the error value was increased since there was such a large difference at the error value: 0. One possible reason for this is that the values of st_3 and st_4 were estimated lower due to the use of data from other observation locations when generating the simulated SNS data.

Table 3 shows that the minimum value is almost 0.00 meters, even when large values are added as errors. However, the maximum value results in an extremely large value. In st_1 and st_2, the mean values of RSME do not change significantly after adding the error value, contrary to this the error value: 0.30 meters in st_4 shows an RSME of 0.52 meters, a large error. Accordingly, this would require a data collection method and analysis process to reduce the error in values, and a process to validate the reliability when large water levels are estimated.

3.2.3 Discussion

We found that as long as the error is small and the ratio of fake data is small, the estimation accuracy does not reduce significantly. All evaluations showed significant improvements compared to the state-space model with observations. This is a result of the fact that the observation locations are about 500 meters away from the estimated locations, whereas the simulated SNS data are within 100 meters, allowing for more accurate estimation. Unlike time-series data, SNS data is sparse in the time direction, although it is effective for estimation if collected at locations that are nearby to the estimated location.

In this evaluation, we also treated the number of SNS data and posted area size as parameters. These two parameters did not significantly affect the estimation results. Hence, we found that even if the number of SNS data is small, locations near the estimation location can be estimated with sufficient accuracy. For st_1 and st_2, the smaller the parameters, the smaller the error. These results indicate that time lag and location gap are allowable within the range of values of this evaluation and that a small number of errors in water level values and fake data prevents a large error. For st_3 and st_4, the errors are large even without adding the water level error, requiring investigation as to the cause. Since this paper covers only four observation locations, we assume that this is due to the effect of low water levels observed at the other locations. Process in 3.1.1 calculated a low accuracy estimation of the observed location. As a result, the simulated SNS data based on the estimation is also calculated lower than the actual water level. Therefore, we need to improve the state-space model itself. One idea is to apply observations and SNS data together to the state-space model and implement a Kalman filter for locations where time-series data is available.

The purpose of our research is to investigate the conditions under which SNS can contribute to flood estimation: how SNS data, if available, can improve the accuracy of flood analysis. In order to achieve the system in Figure **??** Phase 1 validated the effectiveness of SNS data in improving the accuracy of flooding assessments as SNS Data Validation. The results of the assessment using simulated SNS data showed that a certain degree of error was allowable and that the accuracy was better than that of estimation using observation data collected over a longer distance. We conclude that further processing improvements are needed, such as removing larger errors by estimation using a combination of observed and SNS data.

4 Conclusion

This study investigated whether SNS data can be used to assess flooding. Because we believe that SNS data can be effective to figure out flood levels even in places where it is difficult to install, operate, and manage observation devices. Although there have been many studies on flood damage detection using SNS, their effectiveness has not been clarified, and the amount and content of data collected are not fixed depending on the case of flood damage.

This paper validates the SNS data using the following procedure based on our state-space model. The purpose of SNS Data Validation is to investigate whether SNS data can contribute to the accuracy of flood assessment and under what conditions SNS data can improve the accuracy. Since we were unable to determine the error rate contained in the SNS data collected at the time of flooding, we generate quantified SNS data from simulations in this paper. To generate the SNS data, we used time-series data collected from observation devices at multiple locations. The results of flood analysis simulation were assimilated using the time-series data to simulate numerical flood levels. Then, we appended errors to the simulated SNS data . Afterward, we regenerated a flood level on the observation location based on the simulated SNS data with errors and examine the accuracy of the errors on the data assimilation accuracy.

Estimation results show that if water levels contained errors, we found that the errors were small and that if the ratio of fake data was small, the estimation accuracy would not be significantly reduced. All evaluations showed significant improvements compared to the state-space model with observations. Unlike time-series data, SNS data is sparse in the time direction, although it is effective for estimation if collected at locations that are nearby to the estimated location. However, some of the sensors have large errors in the process of calculating simulated SNS data. This error can be derived from the fact that waterway and sewer data were not used in the state-space model. Since waterway and sewer data are difficult to collect, we plan to resolve this issue statistically using a Kalman filter. After a further improvement in accuracy, we will develop and implement Phase 2: Flood Assessment & Promotion Requirement and Phase 3: Flood Assessment.

Acknowledgement

This work was supported by JST, PRESTO Grant Number JPMJPR2036, Japan.

REFERENCES

[1] Intergovernmental Panel on Climate Change Fifth Assessment Report (AR5), Retrieved April 17, 2018, from https://www.ipcc.ch/report/ar5/, 2018.

[2] Milly, P. Christopher D., Wetherald, Richard T., Dunne, K. A., Delworth, Thomas L, Increasing Risk of Great Floods in a Changing Climate, Nature, Nature Publishing Group, Vol.415, No.6871, page514, 2002.

[3] Hirabayashi, Y., Mahendran, R., Koirala, S., and Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., and Kanae, S., Global Flood Risk under Climate Change, Nature Climate Change, Nature Publishing Group, Vol.3, No.9, page816, 2013.

[4] Ernst, J., Dewals, B. J., Detrembleur, S., Archambeau, P., Erpicum, S., Pirotton, M., Micro-scale Flood Risk Analysis based on Detailed 2D Hydraulic Modelling and High Resolution Geographic Data, Natural Hazards, Springer, Vol.55, No.2, pp.181–209, 2010.

[5] Saudi, A. S. M., Ridzuan, I. S. D., Balakrishnan, A., Azid, A., Shukor, D. M. A., Rizman, Z. I., New Flood Risk Index in Tropical Area Generated by using SPC Technique, Journal of Fundamental and Applied Sciences, Vol.9, No.4S, pp.828–850, 2017.

[6] Silva, S. F., Martinho, M., Capitão, R., Reis, T., Fortes, C. J., Ferreira, J. C., An Index-based Method for Coastal-flood Risk Assessment in Low-lying Areas (Costa de Caparica, Portugal), Ocean & Coastal Management, Elsevier, Vol.144, pp.90–104, 2017.

[7] Burn, D. H., Perceptions of Flood Risk: A Case Study of the Red River Flood of 1997, Water Resources Research, Wiley Online Library, Vol.35, No.11, pp.3451–3458, 1999.

[8] Hall, J. W., Meadowcroft, I. C., Sayers, P. B., Bramley, M. E., Integrated Flood Risk Management in England and Wales, Natural Hazards Review, American Society of Civil Engineers, Vol.4, No.3, pp.126–135, 2003.

[9] Miceli, R., Sotgiu, I., Settanni, M., Disaster Preparedness and Perception of Flood Risk: A Study in an Alpine Valley in Italy, Journal of Environmental Psychology, Elsevier, Vol.28, No.2, pp.164–173, 2008.

[10] Pulvirentia, L., Chinib, M., Pierdiccaa, N., Guerrieroc, L., Ferrazzolic, P., Flood Monitoring using Multi-temporal COSMO-SkyMed Data: Image Segmentation and Signature Interpretation, Remote Sensing of Environment, Vol.115, No.4, pp.990-1002, 2011.

[11] Basha,E. A., Ravela, S., Rus, D., Model-based Monitoring for Early Warning Flood Detection, In Proceedings of the 6th ACM conference on Embedded network sensor systems (SenSys), pp.295-308, 2008.

[12] Elshorbagy, A., Corzo, G., Srinivasulu, S., Solomatine, D., Experimental Investigation of the Predictive Capabilities of Data Driven Modeling Techniques in Hydrology—part 2: Application, Hydrology and Earth System Sciences, Vol.14, pp.1943-1961, 2010.

[13] Rafieeinasab, A., Norouzi, A., Kim, S., Habibi, H., Nazari, B., Seo, D., Lee, H., Cosgrove, B., Cui, Z., Toward High-resolution Flash Flood Prediction in Large Urban Areas – Analysis of Sensitivity to Spatiotemporal Resolution of Rainfall Input and Hydrologic Modeling, Journal of Hydrology, Vol.531, part 2, pp.370-388, 2015.

[14] Ruslan,F. A., Samad, A. M., Zain, Z. M., Adnan, R., Flood Prediction using NARX Neural Network and EKF Prediction Technique: A Comparative Study, In Proceeding of the IEEE 3rd International Conference on System Engineering and Technology (ICSET), pp.203-208, 2013.

[15] Sinnakaudan, S. K., Ghani, A. A., Ahmad, M. S. S., Zakaria, N. A., Flood Risk Mapping for Pari River Incorporating Sediment Transport, Environmental Modelling & Software, Elsevier, Vol.18, No.2, pp.119–130, 2003.

[16] Lyu, H. M., Sun, W. J., Shen, S. L., Arulrajah, A., Flood

Risk Assessment in Metro Systems of Mega-cities using a GIS-based Modeling Approach, Science of the Total Environment, Elsevier, Vol.626, pp.1012–1025, 2018.

[17] Hunter, N. M., Bates, P. D., Horritt, M. S., Wilson, M. D., Simple Spatially-distributed Models for Predicting Flood Inundation: A Review, Geomorphology, Elsevier, Vol.90, No.3-4, pp.208–225, 2007.

[18] Horritt, M. S., Bates, P. D., Evaluation of 1D and 2D Numerical Models for Predicting River Flood Inundation, Journal of Hydrology, Elsevier, Vol.268, No.1-4, pp.87–99, 2002.

[19] Kim, J., Hastak, M., Social Network Analysis: Characteristics of Online Social Networks after a Disaster, International Journal of Information Management, Vol.38, No.1, pp.86–96, 2018.

[20] Sufi, F. K., Khalil, I., Automated Disaster Monitoring From Social Media Posts Using AI-Based Location Intelligence and Sentiment Analysis, IEEE Transactions on Computational Social Systems, pp.1–11, 2022.

[21] Teodorescu, H. N., Emergency-Related, Social Network Time Series: Description and Analysis, Time Series Analysis and Forecasting, Springer International Publishing, pp.205–215, 2016.

[22] Hiroi, K., Murakami, D., Kurata, K., Tashiro, T., Shinoda, Y., A Proposal of Data Assimilation Approach for Flood Level Estimation and Evaluation with Urban Flood Disasters, Journal of Information Processing: Consumer Devices and Systems, Vol.10, No.3, pp.55–64, 2020 (In Japanese).

Actual Condition of Disaster Prevention Radio Broadcasting in Atsugi City and its Artificial Intelligence-Based Discrimination of Difficulty in Listening

Takumi Hashimoto*, Takahiro Miura**, Hideo Kasuga*,
Tetsuo Tanaka*, Yoshimichi Ogawa* and Mari Ueda*

*Kanagawa Institute of Technology, Japan
**National Institute of Advanced Industrial Science and Technology, Japan
s2285004@cco.kanagawa-it.ac.jp

Abstract - Atsugi City and Kanagawa Institute of Technology have been discussing how to ensure that disaster prevention radio broadcasts can be heard even in times of disaster, in light of the current nationwide inaudibility of disaster prevention radio broadcasts. Therefore, we conducted acoustic analysis and listening experiments to investigate the actual situation of inaudibility of disaster prevention radio broadcasts in Atsugi City, and attempted to determine whether Artificial Intelligence can automatically discriminate inaudibility. In this paper, we conduct a listening experiment using recorded data of test disaster prevention radio broadcasts and analyze the inaudibility of sound sources for five types of noise and weather conditions. In addition, Convolutional Neural Network is implemented to discriminate the inaudibility of test disaster prevention radio broadcasts by supervised learning image classification using 800 spectrogram images of the recorded data. As a result, Atsugi City's disaster prevention radio broadcasts are considered unlikely to function during disasters caused by typhoons or torrential rains. In addition, it was found that it is difficult to determine the difficulty of hearing disaster prevention radio broadcasts by artificial intelligence in this annotation.

Keywords: disaster prevention radio broadcasting, Artificial Intelligence, Convolutional Neural Network, noise, weather

1 INTRODUCTION

In recent years, disaster prevention information has become indispensable in Japan, where natural disasters have occurred in many parts of the country, and Atsugi City and Kanagawa Institute of Technology's Research Center for Regional Cooperation and Disaster Care have been discussing how information should be provided in times of disaster. In Atsugi City, information on evacuation actions in the event of natural disasters is provided by means of disaster prevention radio broadcasts, e-mail newsletters, and radio broadcasts[1]. However, in view of the current situation in which disaster prevention radio broadcasts are not heard nationwide, it was decided to discuss the possibility of ensuring that disaster prevention radio broadcasts can be heard even when people are outdoors during a disaster.

In this study, we report on acoustic measurement and listening experiments of test disaster prevention radio broadcasts, and automatic identification of inaudibility using AI, with the aim of clarifying the actual condition of inaudibility of disaster prevention radio broadcasts in Atsugi City.

2 HEARING SURVEY FOR THE OGINO AREA

2.1 Summary

Based on discussions with the Crisis Management Division of Atsugi City and the Regional Collaboration Center, we decided on a series of procedures for evaluating the hearing in the Ogino district of Atsugi City, referring to the first edition of "ASJ Technical Standards for Ensuring Performance of Outdoor Loudspeaker Systems in Emergency Situations such as Disasters"[2]. The specific flow was as follows: 1. decision to broadcast a test broadcast for the Ogino area (Atsugi City), 2. decision of the survey date, 3. preparation and distribution of flyers describing the survey to residents in the Ogino area, 4. evaluation of audibility by residents in the Ogino area (Web), 5. analysis and sharing of the results (Kanagawa Institute of Technology). The listening survey was conducted targeting 7,500 people in the Ogino area. On the day of the survey, residents were asked to listen to the test broadcasts outdoors and evaluate the difficulty of hearing the test broadcasts in the form of a questionnaire using a dedicated Web site. The dates and times of the test broadcasts were November 18 (Thursday) and November 21 (Sunday), 2021, at 11:00 a.m.

Table 1: Heraring survey

Total for 2 days	Number of cases	Proportion
All audible	195	42.3%
A little inaudible	125	27.1%
Almost inaudible	95	20.6%
Not at all inaudible	46	10.0%
Total	461	100.0%

Table 2: Heraring survey (Including location data)

Total for 2 days	Tobio	Kamiogino	Nakaogino	Shimoogino	Miharuno	Matsukagedai	Other	Total
All audible	41	36	6	13	25	3	2	126
A little inaudible	31	12	4	3	2	14	1	67
Almost inaudible	10	7	3	42	8	0	5	75
Not at all inaudible	8	1	0	18	0	0	4	31
Total	90	56	13	76	35	17	12	299

2.2 Results

Tables 1 and 2 shows the total results of the two-day hearing evaluation in Ogino area. The evaluation was made on the following four levels: 1. All audible, 2. A little inaudible, 3. Almost inaudible, and 4. Not at all inaudible. The results of this survey showed that the largest number of respondents were able to understand the content of the disaster prevention administrative broadcasts; however, the results of this survey showed that there were some points where many respondents were unable to understand or almost unable to understand the content of the broadcasts.

3 HEARING OF DISASTER PREVENTION RADIO BROADCASTS FOR TESTING

3.1 Acoustic Measurement and Subjective Evaluation

In accordance with the test method[2] described on page 18 on the first edition of the "ASJ Technical Standard for Ensuring the Performance of Outdoor Loudspeaker Systems in Emergency Situations," fixed-point observations of the sound source emitted from the test disaster prevention radio broadcast were conducted at one of the Ogino slave stations. Observations were made for 10 days each at a distance of about 100 to 250 m concentrically from the sound source (Fig. 1). At the time of writing this paper, about 200 days of data were recorded by a PCM recorder (Roland R-07, sampling frequency: 96 kHz, quantization bit: 16 bit). Although the literature [2] stipulates that "situations in which weather has a significant influence on listening tests, such as rain, snow, and strong winds, should be avoided as much as possible", the same method was used regardless of weather conditions in order to compare listening evaluation results in sunny and bad weather conditions.

Fig. 2 shows a spectrogram of the recorded sound sources in rainy weather. The subjective audibility of the observers was evaluated at two levels, A > B. The data evaluated as "A" showed wind noise, but the sound source was not so difficult to hear because it was raining lightly. The data evaluated as "B" were difficult to hear because it was raining so hard that an umbrella was necessary, and the sound of splashing water when a car drove through a puddle masked the sound of the sound source.

3.2 Hearing evaluation in bad weather

Subjective evaluation of listening difficulty was conducted on 11 students (11 males, mean age 21.5 years) using 20 sound source data of test disaster prevention radio broadcasts, including those recorded in the previous study [3]. The headphones used were Beoplay H9i (Bang & Olufsen). A four-point scale was used to evaluate listening difficulty: 1. not difficult to listen, 2. somewhat difficult to listen, 3. quite difficult to listen, and 4. very difficult to listen [4]. Then, ordinal logistic regression analysis was used to examine the effects of five factors (weather (sunny or rainy), presence of passing cars or bicycles, presence of insects, presence of birdsong, and presence of wind noise) on the difficulty of listening. Two types of analyses were conducted: one in which only the main effect was analyzed, and the other in which the main effect and up to first-order interactions were analyzed.

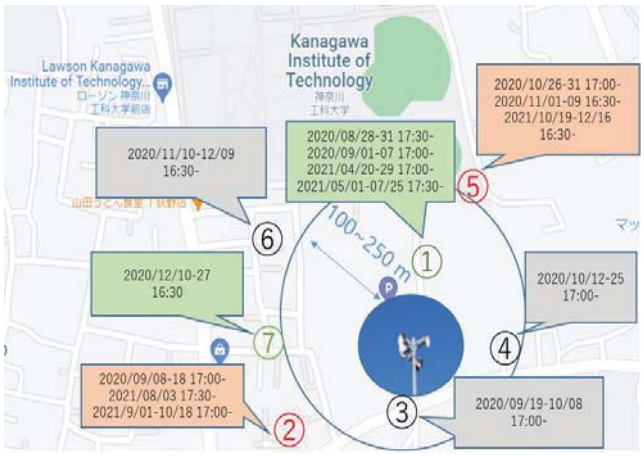


Fig.1 Measurement points and conditions

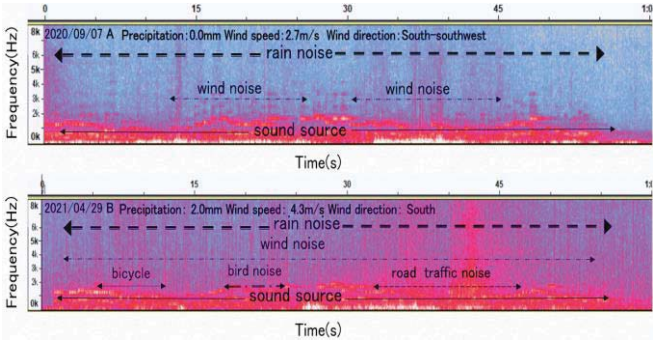


Fig.2 Spectrogram of administrative radio system

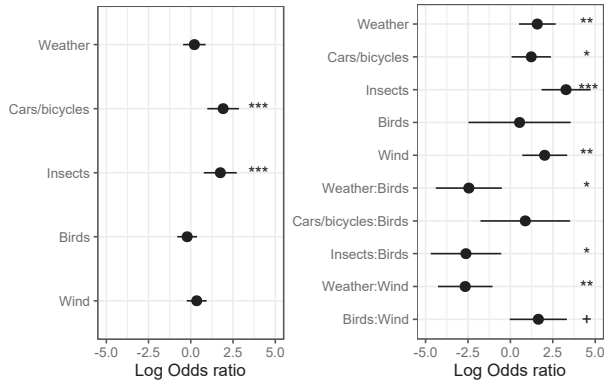


Fig.3 Log odds ratios to listening difficulty on five factors derived by considering only main effects (left) and main effects with two-way interactions (right)

Fig. 3 shows the effect of each factor on the difficulty of listening (+: $p<0.10$, *: $p<0.05$, **: $p<0.01$, ***: $p<0.001$). When analyzing the effect of only the main effect as shown in the left figure of Fig. 3, no significant main effect was confirmed for differences in weather ($p>0.10$). On the other hand, when the effects up to the first-order interaction were checked as shown in the right figure of Fig. 3, a significant main effect was confirmed for the difference in weather ($p<0.01$), and significant interactions were also observed for the difference in weather and the presence of birdsong and the difference in weather and the presence of wind noise ($p<0.01$).

The reason for this result is that, although sound sources should be difficult to hear in rainy weather, when birdsong and wind sound are heard, the effect of rain sound is relatively small due to low precipitation. Therefore, it is possible that the difficulty of hearing the sound was reduced even when it was raining. Conversely, when precipitation is extremely high, birdsong and wind sounds are considered to be less audible.

Based on the above, In the case of earthquake disasters, the audibility will depend on the amount of traffic at the listening point when the weather is clear, and in the case of rain, it will depend on the amount of precipitation, but there is a high possibility that the sound will be inaudible. In addition, in the event of a disaster caused by heavy rain or a typhoon, the sound source will be almost inaudible due to rain noise, and thus the disaster prevention radio broadcasts of Atsugi City are unlikely to function as a means of communicating information.

4 AUTOMATIC IDENTIFICATION OF INAUDIBILITY BY CNN

We implemented Convolutional Neural Network, which is used in the field of image recognition such as image classification and object detection, and created a model that classifies sound sources into two classes, Class A. "easy-to-hear" and Class B. "hard-to-hear", using 800 spectrogram images of test disaster prevention radio broadcast data. Then, we examined whether the accuracy of the model can be used to evaluate the inaudibility of sound sources in the same way as humans do.

4.1 Learning Procedure

The pre-recorded sound sources were rated in four levels (1. not difficult to listen, 2. somewhat difficult to listen, 3. very difficult to listen, and 4. very difficult to listen) as supervised learning, and the correct answers were labeled as 1 is easy to listen and 2 to 4 are difficult to listen, thus discriminating them as a two-class classification problem. The data was classified as a two-class classification problem. In addition, the labeling process is ambiguous in that the noise is not always present during the one-minute sound recording, and it is not clear whether to judge the entire recording as difficult to listen to or only the part with the noise as difficult to listen to. Therefore, we divided the entire one-minute recording data into five-second segments to allow for more accurate labeling and more data. Note that all image data were resized to 143*143 squares.

4.2 Results

Fig.4 shows that the loss values are large and the percentage of correct responses is significantly low, around 50%. Furthermore, the graph is stagnant from the second epoch, suggesting that unlearning is occurring. From Fig. 5, the classification model created was used to classify 40 test data (20 spectrogram images of Class A and 20 spectrogram images of Class B) that are not training data, resulting accuracy of 0.53, precision of 0.53, recall of 0.53, F-Measure of 0.53, and this learning The results showed that the model was not able to discriminate the inaudibility of disaster prevention radio broadcasts in the same way as humans. This result indicates that this learning model could not discriminate inaudibility of disaster prevention radio broadcasts in the same way as humans.

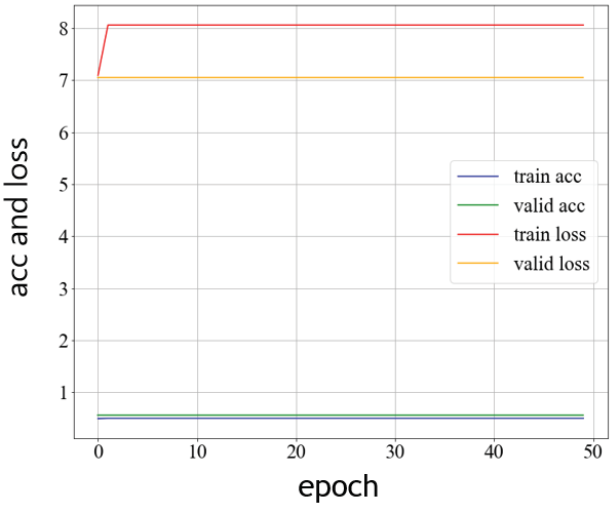


Fig.4 Learning curve with 800 data

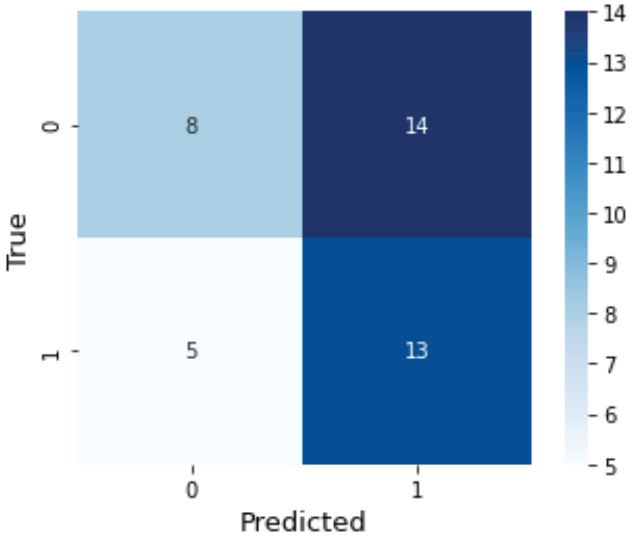


Fig.5 Confusion Matrix

5 DISCUSSION

From the listening survey conducted in the Ogino area, the points where the respondents answered that they could hardly hear or could not hear at all were often located at the edge of residential areas, where the average noise level was high, the distance to the loudspeaker was long, and there were tall buildings or obstacles in the vicinity. These are considered to have a strong influence on the audibility of the disaster prevention administrative broadcasts. However, there are also some stations that are located too close to the disaster prevention radio loudspeaker, but are evaluated to be difficult to listen to. This is thought to be due to the fact that the broadcast itself is audible, but the sound is heard as if it were cracked loudly because it is too close. The failure of automatic identification of inaudibility by the Convolutional Neural Network may well be due to the inadequacy of the model, but it is also possible that the correct label given by subjective evaluation may not always be correct, since the inaudibility of the same recorded data varies from person to person. However, it is also possible that the model is inadequate. In other words, the annotation was not accurate. Therefore, it is necessary to define whether or not it is difficult to listen to each data by multiple people and to assign the correct label to each data.

6 CONCLUTION

From the results, it is clear that Atsugi City's disaster prevention radio broadcasts do not function well as a means of information transmission during disasters caused by rainfall such as typhoons and torrential rains. In addition, it is considered that it is difficult to evaluate the difficulty of listening to disaster prevention radio broadcasts using AI in this annotation. In this listening survey, we received several comments from residents that the way they hear the broadcasts changes depending on the wind direction. In the future, we will investigate how wind speed and direction affect the way of hearing disaster prevention administrative broadcasts. We plan to discuss with Atsugi City the future of disaster prevention administrative broadcasts and the introduction of new information transmission methods using ICT, etc., based on the results of this survey and the fact that we were able to roughly identify points in the Ogino area where it is difficult to hear.

REFERENCES

[1] Disaster Information Transfer / Atsugi City, https://www.city.atsugi.kanagawa.jp/soshiki/kikikanrika/7/8/2310.html [Reference date April 25th, 2022]
[2] ASJ Technical Standard for Ensuring Performance of Outdoor Loudspeaker Systems in Disasters and Other Emergencies,1stEdition, https://asj-msscom.acoustics.jp/?page_id=277
[3] Terashima et al., Acoustical Science and Technology, (Spring), 2020.
[4] Sato et al., Acoustic Journal, 63(5):275-280, 2007.

Session 5:
Social Network and Data Analysis
(Chair: Kozo Okano)

Verification of the Best Time to See Cherry Blossoms Estimation Method Using Time Series Prediction Method

Junya Sato[*], Yusuke Takamori[*], Masahiro Fujimoto[*], Tomonari Horikawa[**], Masaki Endo[***], Shigeyoshi Ohno[***], Masaharu Hirota[****], and Hiroshi Ishikawa[*****]

[*] Electronic Information Course, Polytechnic University, Japan
{b19308, b19309, b19315}@uitec.ac.jp

[**] Nippon COMSYS Corporation
horikawa-tomonari@comsys.co.jp

[***] Division of Core, Polytechnic University, Japan
{endou, ohno}@uitec.ac.jp

[****] Department of Information Science and Engineering, Okayama University of Science, Japan
hirota@ous.ac.jp

[*****] Graduate School of Systems Design, Tokyo Metropolitan University, Japan
ishikawa-hiroshi@tmu.ac.jp

Abstract – With the recent spread of social networking services (SNSs), information from many tourists has come to be shared and collected in real time. While proposing a method for estimating the best time to view cherry blossoms using Twitter, we are advancing research that is expected to help people obtain tourist information. For estimation of the best time to view cherry blossoms, a low-cost moving average method using geotagged tweets was proposed from an earlier study. By combining geotagged tweets and the season-estimation conditions, a person can estimate the best time to see cherry blossoms in any prefecture in Japan. We also proposed a time-series prediction method using machine learning to estimate, for a certain future period, the best time to view cherry blossoms. This method uses past tweet information as training data to predict the number of tweets about cherry blossoms within a certain period. Moreover, the method estimates the best time to view cherry blossoms based on the results. Described herein are the results of estimating the best times to view cherry blossoms for 2022, with confirmation of the method's usefulness.[1]

Keywords: Cherry blossoms; machine learning; SNS; text mining; tourism

1 INTRODUCTION

In today's information-oriented world, many people obtain information from the web. In recent years, with the spread of social networking services (SNSs), a great deal of widely diverse information is posted and viewed by many SNS users. It is not uncommon for tourists use SNSs to gather tourist information when choosing places they would like to visit. Twitter [1] is an example of a social networking service through which tourism information is shared. Twitter is generally used to post and view information about personal events and hobbies. Posting text, photos, and other information related to Twitter is called "tweeting." Among those tweets are some with location information added to a setting, called "geotagged tweets." Geotagged tweets can reflect real-world situations because they can share what happened, where and when. Therefore, they are expected to serve as a social sensor for tourists to estimate and obtain local tourism information in real time. The use of SNS will enable low-cost, seasonal plant seasonality estimation. Currently, biological regular observations are conducted by the Japan Meteorological Agency based on the Biological Seasonal Observation Guidelines. However, seasonal biological observations that are fixed-point observations using sample trees incur costs because of manual observations. The estimation of seasonality using SNS is expected to be useful as a new observation method when biological regulation observation becomes difficult for budgetary reasons. Currently, information related to geotagged tweets on Twitter is used to estimate the best time to see cherry blossoms. Endo et al. [2] proposed a method that uses simple moving averages to estimate the best time to see cherry blossoms. This method, which collects geotagged tweets that include specific key words, estimates the best time to see species by obtaining a simple moving average of the number of tweets per date. In other words, it is possible to estimate the best time to view cherry blossoms in prefectures and localities where a certain number of geotagged tweets are displayed. Using this approach, they confirm the real-world seasonality of tweets related to the names of organisms in each region and demonstrate the potential for presenting tourism information in real time. Takahashi et al. [3] proposed a method using weighted moving averages to improve the estimation accuracy of the method presented by Endo et al. Compared to the method using a simple moving average, this method can improve the results of estimating

This work was supported by JSPS KAKENHI Grant Number JP19K20418, JP20K12081, JP22K13776 and the Okawa Foundation Research Grant.

cherry blossom viewing season for each prefecture while maintaining low costs. Horikawa et al. [4][5] applied the estimation method of Takahashi et al. and proposed a method for estimating the best time to see cherry blossoms in the future using time-series forecasts based on machine learning. Time-series forecasting is the process of predicting data trends over a future period based on past time-series data. Because cherry blossoms are seasonal plants, time-series forecasting is useful to predict the number of tweets related to cherry blossoms for a certain future period. The judgment of when the cherry blossoms are at their best is made by analyzing the relative trends in the number of tweets. Therefore, we have developed a highly accurate estimation of the best time to view cherry blossoms using the time-series predictions of tweets. As described herein, we improve the estimation method presented by Horikawa et al. and newly add tweets for 2022 to estimate the best time to view cherry blossoms. In an earlier study, the accuracy of the estimated cherry blossom viewing time was less than 50% for all the cities tested, thereby rendering accurate estimation impossible. We attributed this finding to the small amount of data used for machine learning. An earlier study used data of 2015–2021. The present study used data through 2022. The accuracy was improved by removing unnecessary data used for machine learning and by improving the data processing methods.

The structure of this paper is the following: Section 2 presents a description of the research related to this paper; Section 3 presents the proposed method; Section 4 explains the experimentally obtained results and their evaluation; and Section 5 summarizes the conclusions of this study.

2 RELATED RESEARCH

Obara et al. [6] use regional associative words and pattern matching from Twitter text to extract tourist information and to elucidate where users live. They have constructed a system that displays the distribution of tweeted locations and estimated user residences by graphing the ratio of acquired regional associative words and the number of users who tweeted tourist information. The method of extracting tourist information using pattern matching achieved a conformance rate of 80.2%.

Silaa et al. [7] propose a new on-the-spot tourist opinion extraction method from the internet for tourism information analysis systems by building a classifier that automatically distinguishes and adds geotagged tweets. They present a method to support lesser-known tourist attractions by automatically collecting and analyzing geotagged tweets and by creating on-the-spot reviews.

Aono et al. [8] proposed a simple method to estimate the flowering date of Someiyoshino cherry trees. The method incorporates data of the temperature during the spontaneous dormancy period. The starting date is obtained from the latitude of the location, distance from the coast, and average temperature during January–March. The model accuracy is improved by estimating the flowering date at 23.8 days, which is the optimal solution common to the entire country for the integrated value. When using the earlier model, 35 of 216 sites had an RMSE of more than 3 days. When using the new model, the number of sites was decreased to 19.

Takamori et al. [9] conducted a forecast of the number of new positive cases based on data published as open data by the Tokyo Metropolitan Government. Using time-series prediction by machine learning and using time-series data including the number of newly positive cases of novel coronavirus and relevant variables as a dataset, they indicate the possibility of using this method to predict the number of newly positive cases in the Tokyo Metropolitan Government.

As described above, some studies have examined extraction of tourist information using SNS, studies of the estimation of cherry blossom viewing time, and studies using time series prediction by machine learning. For the present study, we demonstrate the usefulness of a method for estimating the best time to see cherry blossoms using machine learning time-series prediction for a certain period in the future.

3 PROPOSED METHOD

This chapter presents descriptions of the data and the method used to estimate the timing of cherry blossom viewing.

3.1 Data preparation

For this study, we used the Twitter Streaming API to collect geotagged tweets from Japan that include location information [10]. Then, using the simple reverse geocoding service [11] of the Ministry of Agriculture, Forestry and Fisheries based on latitude and longitude information, the collected geotagged tweets were assigned to the prefecture from which they originated. The timing of cherry blossom viewing was estimated by predicting and analyzing the transition of tweets including specific key words from tweets in each prefecture. The target regions for the experiment were Tokyo, Kyoto, and Shizuoka, all of which have large populations and which are famous for cherry blossoms. The key word was "cherry blossom" in kanji, hiragana, or katakana; the analysis period was February 1, 2015 – February 28, 2022.

3.2 Time series forecast

Based on tweet information collected in the past, a time-series prediction of the number of tweets during the estimated cherry blossom viewing period was performed. Because cherry blossoms bloom mainly during March–April, the estimated period of cherry blossom viewing is the two months of March 1 – April 30, 2022. By performing time-series forecasting of the number of tweets during this period, we can estimate the best seasons for viewing in two-month increments. This method uses Amazon Forecast [12], a service that performs time-series forecasting using statistical and machine learning algorithms from Amazon Web Services [13] (AWS). The training data for this method constitute a large dataset of more than 300 days. In addition, the training data are regarded as seasonal because tweets related to cherry blossoms blooming around spring of each year are collected. Therefore, the Amazon Forecast algorithm uses Prophet, which is suitable for seasonal training data, and DeepAR+, which is suitable for large datasets of 300 days or more. The

predicted value is the value that satisfies the 50% demand obtained from the weighted quantile loss. Weighted quantile loss is a type of metric used during forecasting with Amazon Forecast. The predicted number of tweets is rounded down to the nearest whole number. If the number of tweets is negative, it is treated as 0. The number of tweets during a certain period predicted by machine learning is used to estimate the best time to view the cherry blossoms. The training data used for estimating the best time to view cherry blossoms are tweets from February 1, 2015 through February 28, 2022. No tweet information was available for 2020 as a result of system limitations. Therefore, for prediction, we used tweet information from February 1, 2015 through December 31, 2019, and from January 1, 2021 through February 28, 2022. The time-series forecasting period for the number of tweets is the two-month period from March 1 through April 30, 2022. Prophet is a time series forecasting algorithm based on an additive model for which the nonlinear trend fits year, week, and daily seasonality. It is most effective for time series with strong seasonal effects and several seasons of historical data [14]. When using Prophet, data for all months of the year are used for forecasting. Deep AR+ uses recurrent neural networks (RNNs) for a supervised learning algorithm for predicting 1D time series. It works best with large datasets [15]. DeepAR+ extracts and uses the period from February 1 to April 30, which is a characteristic period for the number of tweets about cherry blossoms. In addition, because the training data for time-series prediction in DeepAR+ use continuous values, the dates of the extracted data are converted so that they become continuous values. For an earlier study, three months of data for each year were processed to fit three months of actual dates when converted to continuous values. Specifically, data for February through April of each year were processed and used as data for November–January, August–October, and May–July; data for February–April were 89 days or 90 days in leap years, which are fewer days than in other seasons. Therefore, when applying the data to another three-month period, the data for May of the same year were used to compensate for a few days of May data to process data for the forecast. In time-series forecasting, cyclical fluctuations in past time-series data are used as a factor in forecasting. Therefore, the data processed in the earlier study might include unnecessary data other than those of February–April, which might therefore impair the data periodicity. Therefore, for this study, to avoid spoiling the periodicity of the time-series data, the data of February–April of each year were converted to continuous values and were used as training data, without supplementing the May data. To give a specific example of processing, we extracted data of February–April from 2015–2021 and February data from 2022, and transformed February 1 through April 30 in 2021 from November 4, 2021 to January 31, 2022, February 1 through April 30 in 2019 from August 7, 2021 to November 3, 2018, and February 1 to April 30, 2018 as May 10 to August 6, 2021, respectively. In an earlier study [5], higher accuracy predictions were obtained when using Prophet.

3.3 Best time estimation method

The condition for estimating the best time to see cherry blossoms by tweet transition uses the conventional method of weighted moving averages. A weighted moving average is a moving average with each value weighted. In the method used for this study, the median value is set to 1. The minimum and maximum values are set respectively as ±0.5 from the median value. To estimate the best viewing period for cherry blossoms, we analyze the frequency of geotagged tweets including the key word. The following three values will be analyzed by date and will be used as the basis for estimating the optimal viewing period.

1. 1-year simple moving average of the number of tweets
2. 7-day weighted moving average of tweets
3. 5-day weighted moving average of tweets

To assess the increase in the frequency of geotagged tweets containing the key words, a 1-year simple moving average of the number of tweets was obtained. Because tweets including key words have words that are unrelated to actual cherry blossoms, one can ascertain the increase in tweets related to actual cherry blossoms by comparing the number of tweets with the one-year simple moving average of the number of tweets. To analyze the increasing trend of the number of tweets, a 7-day weighted moving average and a 5-day weighted moving average are obtained. Because the number of tweets tends to increase on Saturdays and Sundays, a 7-day weighted moving average is obtained. The formula to be used is shown below. x_y denotes the number of tweets x of y days prior.

$$H_{avg7} = \frac{0.5x_7+0.67x_6+0.83x_5+x_4+1.17x_3+1.33x_2+1.5x_1}{7} \quad (1)$$

One can also find the weighted moving average of 5 days, which is the average number of days between the cherry blossom bloom and full bloom. The formula to be used is presented below. x_y is the number of tweets x of y days prior.

$$H_{avg5} = \frac{0.5x_5+0.75x_4+x_3+1.25x_2+1.5x_1}{5} \quad (2)$$

Using these estimation criteria, we can ascertain that cherry blossoms are at their best when the following two conditions are met.

1. Number of tweets is greater than the 1-year simple moving average
2. The 5-day weighted moving average of the number of tweets is greater than the 7-day weighted moving average for three consecutive days

These conditions are used to estimate the best time period to see each attraction.

4 EXPERIMENT RESULTS

In this chapter, we present results of the estimation of cherry blossom viewing time using this method. The peak cherry blossom peak prediction can be achieved by combining tweet information with the best time-of-view estimation conditions. The accuracy of the peak prediction is

evaluated by comparing results of peak estimation with the correct prediction period, and by the values of the reproduction rate and the goodness-of-fit rate. The correct prediction period is defined as the period from the cherry blossom blooming date to the full blooming date, as observed by the Japan Meteorological Agency [16]. The cherry blossom date represents the first day when five to six or more cherry blossoms are in bloom on a specimen tree. A cherry blossom full bloom date is the first day on which 80% or more of the specimen trees are open. First, Figures 1–3 respectively present results for Tokyo, Kyoto, and Shizuoka prefectures for 2022: estimates of the cherry blossom viewing season based on the number of tweets obtained from the time-series forecast.

Figures 4–6 respectively present estimated results of the best time to see cherry blossoms in Tokyo, Kyoto, and Shizuoka prefectures in 2022, based on the number of tweets obtained using DeepAR+ as the algorithm for time-series prediction.

Figures 7–9 respectively present results obtained from estimating the cherry blossom viewing season in Tokyo, Kyoto, and Shizuoka prefectures in 2022. Table 1 presents the actual number of tweets and the predicted number of tweets. Table 2 shows the reproducibility and goodness-of-fit of the estimation of the cherry blossom season using the actual number of tweets and the predicted number of tweets.

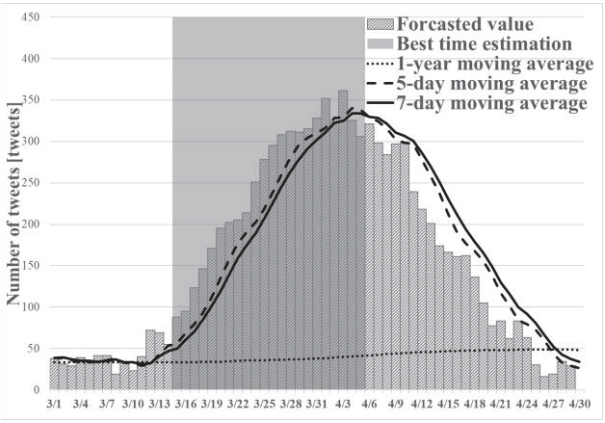


Figure 1: Tokyo after forecasting in 2022 (Prophet).

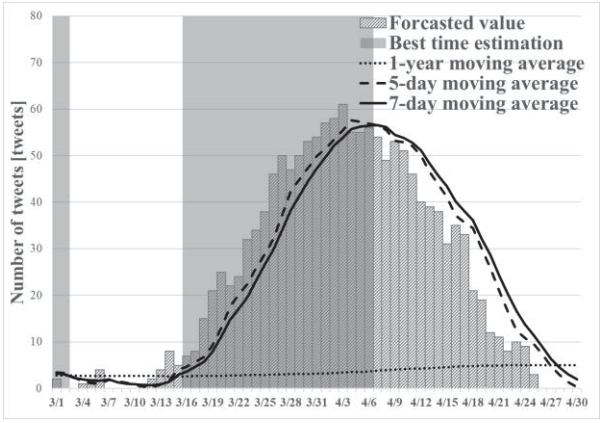


Figure 2: Kyoto after forecasting in 2022 (Prophet).

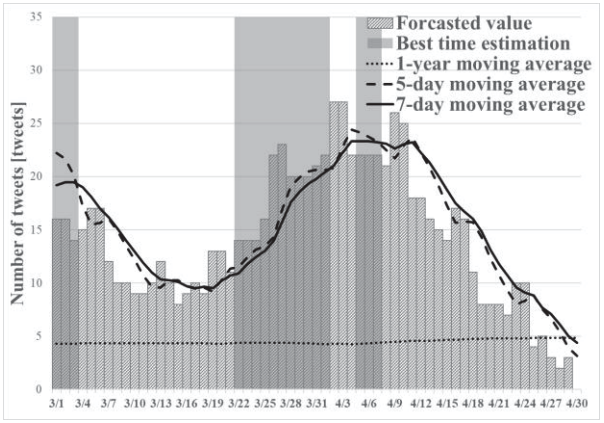


Figure 3: Shizuoka after forecasting in 2022 (Prophet).

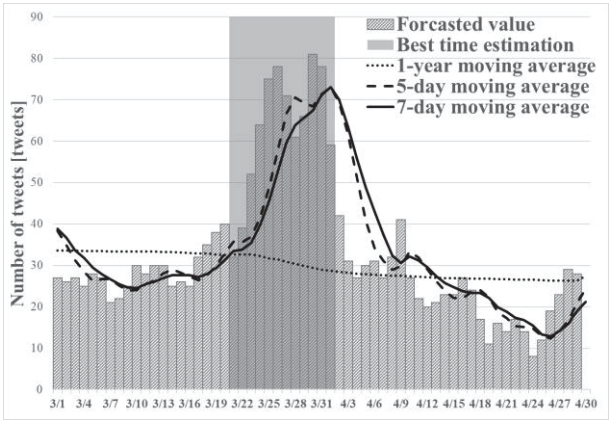


Figure 4: Tokyo after forecasting in 2022 (DeepAR+).

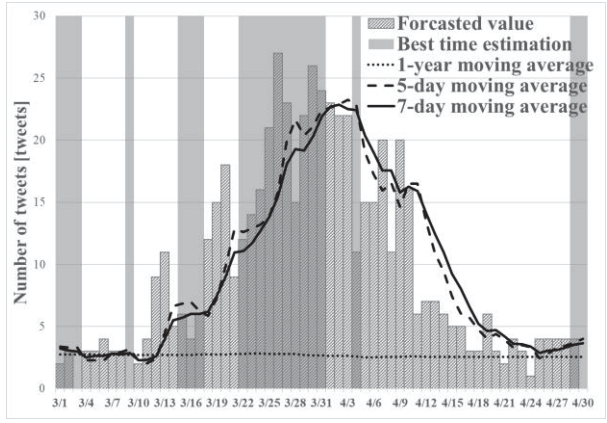


Figure 5: Kyoto after forecasting in 2022 (DeepAR+).

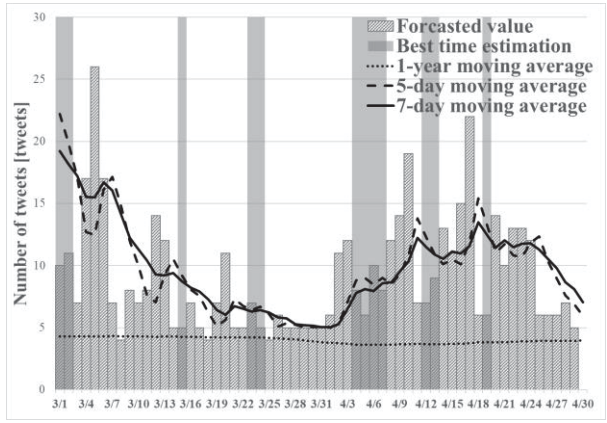


Figure 6: Shizuoka after forecasting in 2022 (DeepAR+).

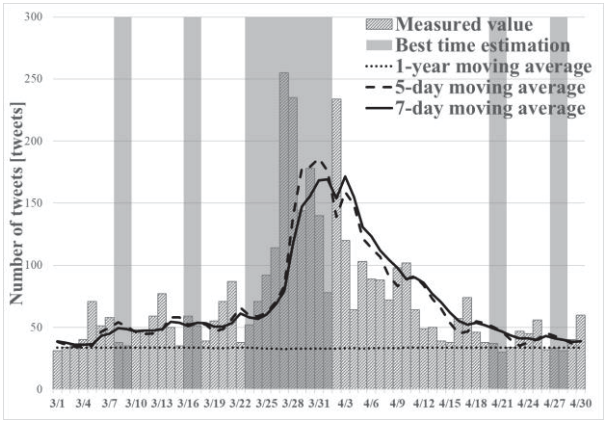


Figure 7: Tokyo in 2022.

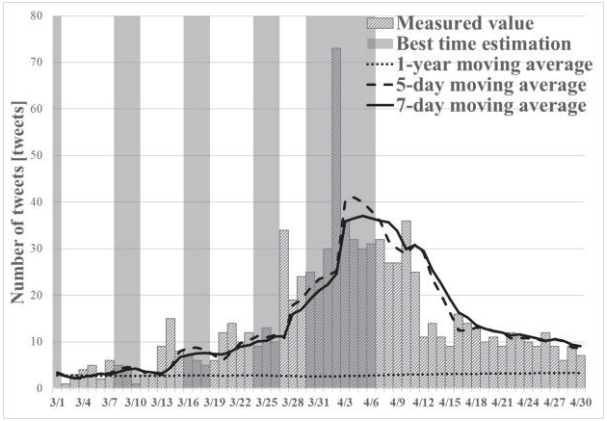


Figure 8: Kyoto in 2022.

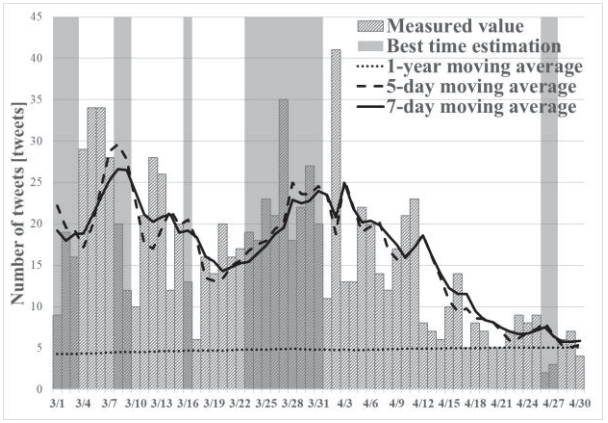


Figure 9: Shizuoka in 2022.

Table 1: Total numbers of actual and expected tweets

	Area	Number of tweets
Actual		4280
Prophet	Tokyo	9624
DeepAR+		1979
Actual		873
Prophet	Kyoto	1459
DeepAR+		578
Actual		956
Prophet	Shizuoka	863
DeepAR+		543

Table 2: Reproducibility and goodness-of-fit rates for actual tweets and predicted tweets for the estimated viewing method

	Area	Recall	Precision
Actual		62.5%	27.8%
Prophet	Tokyo	100.0%	36.4%
DeepAR+		87.5%	58.3%
Actual		57.1%	22.2%
Prophet	Kyoto	100.0%	29.2%
DeepAR+		100.0%	35.0%
Actual		80.0%	47.1%
Prophet	Shizuoka	90.0%	52.9%
DeepAR+		20.0%	16.7%

A comparison of Figures 1–9 is illustrative. Comparison of Figures 1–3 and 4–6 with Figures 7–9 confirms the predictability of the trend in the number of tweets for both algorithms. Particularly, only in the trend of the number of tweets in Shizuoka prefecture can we confirm the existence of a peak in early March, both in terms of actual and predicted number of tweets. This finding is attributable to the fact that Shizuoka prefecture is famous for its early blooming cherry trees such as Kawazu-zakura (in Japanese 河津桜) and Kakegawa-zakura (in Japanese 掛川桜). As an example, "Kawazu Cherry Blossoms are in full bloom! Kawazu-zakura, the original cherry tree, known throughout Japan as an early blooming cherry tree, is now in full bloom." the tweet reads. In fact, 108 of the 211 tweets posted during March 1–10 included the words "Kawazu-zakura or "Kawazu-gawa-zakura-namiki." Therefore, it is conceivable that the trend of tweets related to Kawazu cherry blossoms was predicted.

Table 1 presents the results. The total number of tweets predicted using Prophet tended to be higher than the actual number of tweets in Tokyo and Kyoto prefectures, but lower

than the actual number of tweets in Shizuoka prefecture. The total tweets predicted by DeepAR+ tended to be fewer than the actual tweets in Tokyo, Kyoto, and Shizuoka prefectures.

Table 2 also presents results. The time-series prediction of the number of tweets using Prophet showed higher reproducibility and better goodness-of-fit than the method using the actual number of tweets. The reproducibility of Tokyo improved from 62.5% to 100%; the fit rate improved from 27.8% to 36.4%. Kyoto prefecture's reproducibility improved from 57.1% to 100%; its compliance rate increased from 22.2% to 29.2%. Shizuoka prefecture's reproducibility improved from 80% to 90%; its compliance rate increased from 47.1% to 52.9%.

For Tokyo, Kyoto, and Shizuoka prefectures, the reproducibility and fit rates were higher than those obtained using the timing estimation method with the actual number of tweets. Shizuoka prefecture in particular achieved a remarkably high reproducibility rate of 52.9%.

The time-series prediction of the number of tweets using DeepAR+ demonstrated the possibility of predicting the number of tweets in Tokyo and Kyoto prefectures with higher reproducibility and goodness-of-fit rates than the method using the actual number of tweets, although it was not possible to predict the number of tweets in Shizuoka prefecture with high reproducibility and goodness-of-fit rates. The reproduction and fit rates of time-series forecasting method using DeepAR+ and the method based on actual tweet counts were compared. The reproduction rate for Tokyo improved from 62.5% to 87.5%; the fit rate improved from 27.8% to 58.3%. Kyoto prefecture's reproducibility improved from 57.1% to 100%; its conformance rate increased from 22.2% to 35.0%. Shizuoka prefecture's reproducibility decreased from 80% to 20%; its compliance rate declined from 47.1% to 16.7%.

5 CONCLUSION

As reported herein, we conducted experiments to assess a time-series forecasting method for estimating the best time to view cherry blossoms at a certain time in the future. The experiment results indicated the usefulness of this time-series forecasting method using machine learning.

When the number of tweets in the predicted time series is used to estimate the best time to view cherry blossoms, the repeatability and goodness-of-fit rates are expected to be as good as or better than those obtained by estimating the best time to view cherry blossoms using actual tweets, except for the time series in Shizuoka prefecture using DeepAR+. Therefore, results confirmed the possibility of predicting the best time to view cherry blossoms for a certain period of time in the future using an appropriate algorithm. Comparison of the results obtained respectively using Prophet and DeepAR+ demonstrated that the accuracy of time-series prediction using DeepAR+ was higher in Tokyo and Kyoto prefectures, although the accuracy of time-series prediction using Prophet was higher in Shizuoka prefecture. An earlier study showed that the time-series forecasting method using DeepAR+ failed to improve the reproducibility and goodness-of-fit rate significantly compared to the method using actual tweets. By contrast, this study showed higher reproducibility than that of the time-series forecasting method using Prophet in Tokyo and Kyoto prefectures.

For this study, we demonstrated the possibility of using a method for estimating cherry blossom season with time-series prediction of the number of tweets to predict cherry blossom season trends for a certain future period. Additionally, we demonstrated the possibility of using both Prophet and DeepAR+ as algorithms for time-series forecasting to achieve higher accuracy. Particularly, the time-series forecasting method using DeepAR+ enabled more accurate forecasts than Prophet by screening data used for time-series forecasting. Additionally, results show that the use of tweet data enables low-cost estimation of cherry blossom seasonality for seasonal plants. In future studies, we intend to increase the number of target areas and to achieve more accurate forecasts using new cherry-blossom-related variables such as temperature.

REFERENCES

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

[2] M. Endo, K. Mitomi, K. Saeki, Y. Ehara, M. Hirota, S. Ohno, and H. Ishikawa, Study of information provided by the best time to see estimation method of phenological observations using tweets, Proceedings of the 12th Annual Conference on Japan Society for Tourism Informatics, Shizuoka, Japan, pp.47-60 (2016).

[3] M. Takahashi, M. Endo, S. Ohno, M. Hirota, and H. Ishikawa, Automatic detection of tourist spots and best-time estimation using social network services, International Workshop on Informatics 2020, pp.65-72 (2020).

[4] T. Horikawa, M. Takahashi, M. Endo, S. Ohno, M. Hirota, and H. Ishikawa, Estimating the best time to see cherry blossoms using SNS and time-series forecasting of tweet numbers using machine learning, International Workshop on Informatics 2021, pp.37-44 (2021).

[5] T. Horikawa, M. Takahashi, M. Endo, S. Ohno, M. Hirota, and H. Ishikawa, Estimating the Best Time to View Cherry Blossoms Using Time-Series Forecasting Method, Multidisciplinary Digital Publishing Institute, pp.418-431 (2022).

[6] M. Ohara, K. Morita, M. Fuketa, and J. AOE, Extraction of Tourist Information from Contents of Tweets and Building an Analysis System, The 29th Annual Conference of the Japanese Society for Artificial Intelligence, Vol.29, pp.1-3 (2015).

[7] V. Silaa, F. Masui, and M. Ptaszynski, Vol.12, No. 5, 2321, pp.1-24 (2022).

[8] Y. Aono, and N. Murakami, A simplified method to estimate cherry blossom phenology considering temperature during endodormancy process, Climate in Biosphere, Vol.17, pp.25-33 (2017).

[9] Y. Takamori, R. Watanabe, D. Kato, M. Endo, and H. Ishikawa, Numerical estimation of new COVID-19 positive cases using time series analysis by machine learning, Proceedings of the 13th International Conference on Management of Digital EcoSystems, pp.153-159 (2021).

[10] Twitter Developers, https://dev.twitter.com/ (referred

in June 2022)
[11] Agricultural Research Institute, Simple Reverse Geocoding Service, https://aginfo.cgk.affrc.go.jp/rgeocode/index.html.ja (referred in June 2022)
[12] Amazon Forecast, https://aws.amazon.com/jp/forecast/ (referred in June 2022)
[13] AWS, https://aws.amazon.com/jp/console/ (referred in June 2022)
[14] Prophet, https://docs.aws.amazon.com/ja_jp/forecast/latest/dg/aws-forecast-recipe-prophet.html (referred in June 2022)
[15] DeepAR+, https://docs.aws.amazon.com/ja_jp/forecast/latest/dg/aws-forecast-recipe-deeparplus.html (referred in June 2022)
[16] Japan Meteorological Agency https://www.data.jma.go.jp/sakura/data/index.html (referred in June 2022)

A Study of Different Values between Individuals and Companies Regarding Personal Information

Mayu Satake, Hideki Goromaru*

*Faculty of Social Systems Science, Chiba Institute of Technology, Japan
goromaru.hideki@p.chibakoudai.jp

Abstract - In recent years, with the spread of mobile devices such as smartphones and tablets, there have been more and more opportunities to easily post text, photos, videos, and other information on social networking services (SNS) on the internet and add comments. In addition, SNSs has made it possible to communicate not only with acquaintances but also with people who have never met before. However, there have been risks that outsider who are unhappy with the content of comments or posting procedures may misuse information posted on SNSs to engage in problematic behavior such as stalking or harassment. To reduce these risks as much as possible, it has been important to avoid posting information that are vulnerable to misuse. However, we do not know what kinds of information are vulnerable to misuse. Therefore, we first conducted investigates focusing on the value of information, assuming that information that are easily misused are information of high value, such as personal information. As a result, we have found that, except for information related to attracting customers such as gender, occupation, and grade, individuals estimate the value of information higher than companies and organizations.

Keywords: Personal Information, Personal Data, Sensitive Information, Privacy Information

1 INTRODUCTION

In recent years, with the spread of smartphones and tablet terminals, people have more opportunities to easily post texts, photos, and videos on SNSs on the Internet. It has become possible to communicate not only with acquaintances but also with people we have never met. And many SNSs allow users to set the public range so that users can receive as little damage as possible from outsiders [1]. However, even so, depending on the content of comments and the way they are posted, outsiders may engage in problematic behavior that users do not want, such as stalking or harassment, and may misuse the information posted on SNSs [2]. In addition, even if SNS's users pay attention to security and privacy on a daily basis, they may become distracted due to poor health or security fatigue and post unintentional content [3]. Therefore, it is important to avoid disclosing information on SNSs that could be misused as much as possible.

However, the act of not disclosing information out of concern for misuse is not an intended use of SNSs and reduces its usefulness. Therefore, it is necessary to maintain the enjoyment of SNSs, such as communication with outsiders, self-disclosure, socializing, information acquisition, and entertainment, while keeping the damage caused by misuse-prone information that should not be posted on SNS within acceptable limits, in other words, to strike a balance between opportunities and threats.

However, it is unknown what kind of information is easily misused. In particular, if there is a discrepancy between the personal value perception of information and the value perception of the world, it may not be possible to properly select due to the discrepancy. For example, information that is considered to be okay for individuals to publish may cause great damage if it is published on SNSs.

Therefore, we focused on the value of information and considered information that is easily misused as information of high value. Next, we considered taking that value as a monetary amount and quantifying it. Then, in order to grasp the difference between individual and general value perceptions of information, we investigate the amount of money for information by questionnaire in the case of individual value perceptions, and by court records of companies and organizations in the case of general value perceptions.

In this paper, the value of information and data is quantified for the goal of reducing the damage caused by information and data related to individuals within the permissible range while maintaining the convenience of SNSs as much as possible. Chapter 2 is the definition of information and data related to individuals, Chapter 3 is the current state of data and information on SNSs, Chapter 4 is the problems and issues of conventional research, and Chapter 5 is the proposal of information related to individuals and the value estimation model of data. Chapter 6 describes the results. Chapter 7 describes the discussion and Chapter 8 describes the conclusions.

2 DEFINITIONS OF PERSONAL DATA

In this section, we organize the terms because there are several terms with the same or similar names. For example, "Personal Data" under Japanese law [4] and "Personal Data" in the European GDPR are different concepts despite having the same name. "Sensitive information" and "Special care-required personal information" are similar concepts but different terms. Data and information are strictly different concepts. However, data and information are sometimes used in the same class. To avoid confusion in this paper, "data and information on individuals" is unified as "Inf" and divided into Inf.1 through Inf.6, and the position of each term is defined in Figure 1 and Table 1. Figure 1 is based on existing literature on Personal Data (under the EU's GDPR), personal information and privacy [5][6][7][8][9][10].

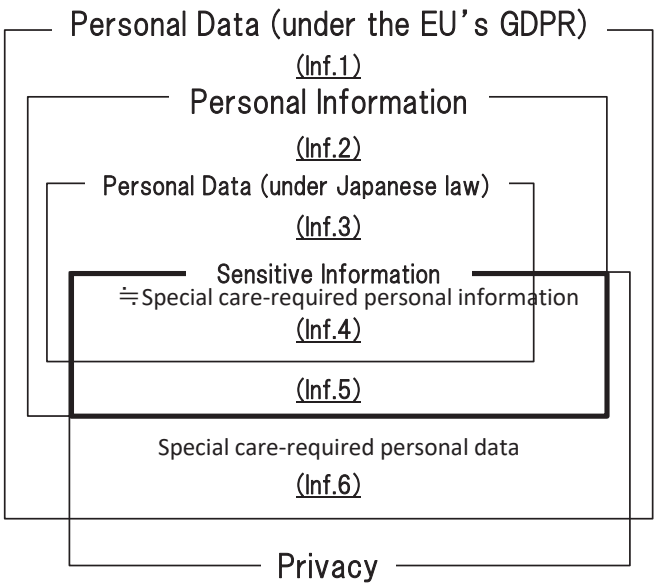


Figure 1 Personal Data and Privacy

Table 1 Features of Inf.

Inf	Individually identifiable	Searchable	non-public
1			
2	✓		
3	✓	✓	
4	✓	✓	✓
5	✓		✓
6			✓

In this paper, we define it as follows:

- "Personal Data" is not Personal Data under Japanese law, but Personal Data under the EU's GDPR. "Personal Information" includes Inf.1, Inf.2, Inf.3, Inf.4, Inf.5 and Inf.6.
- "Personal Information" includes Inf.2, Inf.3, Inf.4 and Inf.5.
- The terms "Sensitive information" and "Special care-required personal information" are unified to "Sensitive information". "Sensitive information" includes Inf.4 and Inf.5.
- "Special care-required personal data" includes Inf.6.

3 DATA AND INFORMATION ON SNS

It is easy to judge whether Inf.1 corresponds to Personal Data. Personal Data that individuals should pay particular attention to when posting on SNSs are Inf.4, Inf.5 and Inf.6 that require careful handling. In the research of Humphreys.L et al. [11], Simple information with personal identification such as e-mail address was posted on SNSs turns out that this is rare.

3.1 Attribute Information (Inf.1) and Identification Information (Inf.2~Inf.5)

Ishii [12] classifies information on SNSs into two categories: identification information and attribute information as follows:

(1) Attribute information (Inf.1)
- Information that cannot be identified by one piece of information (example: hobbies, occupation / grade, family structure, gender, current location, movement history, age, birthplace, address to prefecture, presence / absence of lover's spouse, blood type, etc.).

(2) Identification information (Inf.2~Inf.5)
- Information that combines basic information (example: name, affiliated company, school, face photo, address, etc.)

In the case of SNSs, the identification information is a factor to expand the friendship to "known friends", but it is not a factor to expand to "friends only by SNSs". On the contrary, attribute information is a factor that expands friendships in both cases. In addition, Sato et al. [13] consider that Personal Information on the Internet can be classified into four categories, identification information is similar to contact information, and attribute information is a hobby. Therefore, when posting to an SNS that can be viewed by an unspecified number of people, it is often posted for the purpose of communicating with an unspecified number of people. In that case, attribute information rather than identification information is easier to disclose self-information.

3.2 Sensitive Information (Inf.4 and Inf.5)

When companies utilize information or data provided or collected from individuals, Sensitive Information (e.g., thought, physical characteristics: Inf.4 and Inf.5) cannot be used without their consent [4]. However, in the case of individuals, the boundary definition of information to be protected is ambiguous and depends on the individual's way of thinking. Machida et al. [14] investigated information leakage after SNSs postings from 137,877,745 postings from June 12, 2012 to June 13, 2013, and found that 0.008% of Inf.4 and Inf.5 existed. Although this detection value is very low, it is necessary to detect even a very small number of Inf.4 and Inf.5 to prevent leaks.

3.3 Special care-required personal data (Inf.6)

In the case of Issues of Special care-required personal data (Inf.6), which requires careful handling, there are two factors that cause it to be posted, although we are careful when posting it to SNSs. The first point is "in the case which it is used for self-disclosure, entertainment, and socializing with friends." The second point is "in the case of human error".

The first point is that there are many cases where an SNS is used for self-disclosure, entertainment, and socializing with friends. Reasons for participating in the "online

community," which can be said to be the predecessor of SNSs, are the four factors of self-disclosure, socializing, information acquisition, and entertainment. Entertainment value, social promotion, and group norms have been found to influence Facebook[1]'s intent to use [15] [16]. In addition, the use of Facebook contributes to the maintenance and formation of social capital for university students [17].

The second point is the case of human error. For example, location information such as GPS is attached to SNSs postings, and the location information function can be turned on or off depending on the situation at the time of posting. However, forgetting to turn it off due to human error can lead to leakage of current location and movement history. In a study by Tehila et al. [18], 81% (2,383) of the surveyed Facebook users posted family-related information such as family name, date of birth, and photo, according to a survey of Facebook users. Given these facts, the possibility of information leakage remains, and caution should be exercised when posting on SNSs.

3.4 Ease to search and ease to change

Ishikawa et al. [19] show two points to consider when Personal Data leaks on SNSs. The first point is "information that can be easy to search". The second point is "information that can be easy to change".

The first point, " information that can be easy to search," is information that makes it easy to connect the person who leaked the information and the person on the SNSs (it turns out that they are the same person) when the information leaks. For example, an email address is information that can identify an individual, is easy to link to another SNSs, and is easy to access, so care must be taken when handling it. On the other hand, if a hobby or profession is not a special hobby or profession, it cannot be easily identified as the person himself / herself, so there is no need to be careful in handling it. It is important to be "difficult to search".

The second point, "information that can be easy to change," refers to information whose contents can be easy to change even if the information is leaked. By changing contents, it is possible to make it impossible to easily identify the person who leaked the information and the person on the SNSs as the same person. For example, the date of birth and family structure cannot be changed, but the email address and mobile phone number can be changed. It is important to be "easy to change".

"Difficulty to search" and "ease to change" are major factors influencing the value of information. Conversely, "ease to search" and "difficulty to change" are major factors that increase the negative effects of information leaks. These are different from the concepts of Inf.1 to Inf.6 of Personal Data. In other words, it is unique concepts of information leakage by SNS's users, which is different from information leakage of companies.

4 PREVIOUS WORK

Here, we describe a conventional model that calculates the amount of money from the value of information.

[1] Facebook is a registered trademark of Facebook, inc.

4.1 Information Value Calculation Model

Ponemon [20] proposes a formula to calculate the average cost per case for the occurrence of personal information leakage. Romanosky et al. [21] have created a cost calculation model formulated from incident information of Advicen in the United States. However, it is a formula that focuses on the corporate cost from the amount of leaked information instead of each personal information, and it is not possible to calculate the price of each one just by knowing the total price.

There are also calculation models such as conjoint analysis [22] and CVM [23]. However, since these methods are analytical methods for investigating the needs of victims, they are inappropriate methods for investigating the value of objective information. Therefore, in order to obtain the value of objective information, we considered to use the results of the trial and the ideas of companies and organizations as the basis. Then, it is necessary to investigate how the users of SNSs perceive the value of personal information and to investigate the difference from the basic axis.

4.2 JO Model

JNSA [24] aggregates information on incident articles reported in newspapers and Internet news since 2002, and documents related to incidents released by organizations, and calculates the estimated amount of damage compensation for each customer of each company. The JO model (JNSA Damage Operation Model for Individual Information Leak) is a method that can calculate the price of each item and is suitable for our purpose because it is designed for companies and organizations to understand the potential risk of personal information. The calculation formula is shown below.

Damage compensation amount
= Leaked information value × Social responsibility degree
× Post-event response degree
= (Basic information value × Sensitive Information degree
×Personal identification degree)
× Social responsibility degree
× Post-event response degree
= Basic information value [500]
×**Sensitivity Information degree** [max ($10^{\max(x)-1} + 5^{\max(y)-1}$)]
× **Personal identification degree** [6, 3, 1]
× **Social responsibility degree** [2, 1]
× Post-event response degree [2, 1]

The main indicators are as follows:

(1) Sensitivity Information degree

- The leaked personal information is calculated using a table showing **"psychological distress"** x (3 values) and **"economic loss"** y (3 values)
- For example, name is x = y = 1, disease name is x = 1 and y = 2, etc.

(2) Personal identification degree

There are three values of judgment criteria.

- If it is possible to easily identify an individual with "name" and "address", the value is "6".
- If there is a possibility that an individual can be identified by adding the cost with "name" or "address + phone number", the value is "3".
- If it is difficult to identify other than the above, the value is "1".

(3) Social responsibility degree

There are two values of judgment criteria.

- In the case of business (medical, financial, credit, etc.) that handles personal information, the value is "2".
- In the case of other general companies, associations, and organizations, the value is "1".

However, the JO model has multiple problems [25]. Regarding the basic information value of the JO model, the constant of 500 yen is determined by the subjectivity of experts, and there is no basis for it. Also, regarding the degree of social responsibility and the degree of accident response, the JO model is an old model designed nearly 20 years ago. The law at the time of the judgment has many differences from the recent law. Also, regarding the coefficient used in the JO model, the coefficient was determined by the subjectivity of the experts at that time and has no basis.

Therefore, Yamada et al. [25] proposed a new mathematical model (here, it is called Yamada-Model). Table 2 shows the coefficients of the new indicators (Economic Rank of Sensitive Information degree, Spiritual Rank of Sensitive Information degree, Ease of Identification) in Yamada-Model. This Yamada-Model made it a more realistic model than the JO model.

Table 2 Comparison of coefficients between JO model and Yamada-Model

Indicators	Comparison	Three-step coefficient		
Psychological distress	JO-Model	5^0	5^1	5^2
	Yamada-Model	1	1.0129	1.0261
Economic loss	JO-Model	10^0	10^1	10^2
	Yamada -Model	1	1.1723	1.3743
Personal identification	JO-Model	1	3	6
	Yamada -Model	1	1.5158	2.8291

Note: Psychological distress and Economic loss are included in "**(1) Sensitivity Information degree**." Personal identification is "**(2) Personal identification degree**."

5 NEW MODEL

From the Yamada-Model, which is an improved version of the JO model, by changing its coefficients and equations, we can estimate the value of the information that a company considers. And these models are a method to determine the impact on victims of personal information leaks based on " psychological distress" and "economic loss". However, it is based on the leakage of personal information of companies, and is different from the leakage of personal information by users of SNSs. Therefore, we create a new model for estimate information values on SNSs .

5.1 Conversion of indicators according to SNS's users

In the case of JO-Model and Yamada-Model, the coefficients were "psychological distress" and "economic loss". However, personal information leakage on SNSs and corporate personal information leakage are different. Therefore, we create a model tailored to personal information on SNSs.

In this paper, we replace "psychological distress" with "difficulty to change" information. "Difficult to change" information refers to information that cannot be changed, such as the date of birth. Once leaked, it will remain on the Internet forever, causing psychological distress. Due to this causal relationship, "difficulty to change" information becomes one of the factors of "psychological distress".

In the same way, we replace "economic loss" with "easy to search" information. "Easy to search" information refers to information that can be easy to identify some persons with search, such as an email address and phone number. Once leaked, the negative effects will be widespread rapidly and the information be exploited, causing economic loss possibly. Due to this causal relationship, "easy to search" information becomes one of the factors of "economic loss".

5.2 Information and Data Classification

Some of the coefficients in the JO-model are subjective or based on outdated data. Therefore, we decided to conduct this survey in order to create objective and up-to-date coefficients while utilizing the Yamada- model. The survey items included a total of 15 items of Sensitive Information: hobbies, occupation and grade, family structure, gender, current location and travel history, landline phone number, age and date of birth, self-portrait, e-mail address, cell phone number, address, school and work name, name, and phone number. These items were picked up from the surveyed Inf.1 to Inf.6 basing on Chapter 3. The reason for this categorization is that if we simply extracted only those items that are frequently posted on social networking sites, most of them would be attribute information (Inf.1), which would lead to bias.

Section 3.4 discussed "ease to search" and "ease to change". These are some of the key concepts involved in information leakage for SNS's users. We considered applying this to the New-Model, which is an improvement of the Yamada-Model, in combination with Section 5.1. The combinations are shown in Table 3. In Table 3, the higher up and to the right, the higher the risk.

Table 3 was created based on the following judgment in order to process qualitative information quantitatively. Table 3 is evaluated on a three-values scale with reference to the Yamada-Model for both "Changeable information" and "Searchable information".

- "Changeable information" conducts a graded evaluation based on the time it takes to make changes. Specifically, Easy (changeable within 2 weeks), Normal (changeable within 2 weeks to 1 year), Difficult (changeable within 1 year or more before change).

・ "Searchable information" conducts a graded evaluation based on the susceptibility to secondary damage. Easy (easy to change), Normal (changeable), Difficult (unchangeable or difficult)

Table 3 Matrix of Searchable information and Changeable information

		Changeable information		
		Easy	Normal	Difficult
Searchable information	Easy	⑦ Email address, cell phone number	⑧ Address, school name / work place, name, family / friend's name	⑨Medical records
	Normal	④ Current location / movement history	⑤ Fixed phone number	⑥ Age, date of birth, own photo
	Difficult	①Hobby	②Occupation / Grade	③ Family structure, gender

5.3 Plot method and Formula Coefficients

This section discusses ①through ⑨in Table 3.

① For example, hobby. Least risk.

② For example, occupation, grade.
- Occupation is difficult to specify except for special occupations, and it is often difficult to change within 2 weeks.
- Also, if only the grade is leaked, it will not be identified, but it is difficult to change within 2 weeks.

③ For example, family structure, gender.
- Family structure and gender are difficult to identify the individual, but unlike hobbies and occupations, it is impossible to change or it takes more than a year to change.

④ For example, current location / movement history.

⑤ For example, fixed phone number.
- Fixed phone numbers are information that is susceptible to secondary damage if leaked. And in order to change the number, it is not permitted to change the number except for prevention of wrong calls or prevention of nuisance calls. Therefore, it is information that takes time to change. However, in 2020, the prevalence rate was 68.1% and has been declining in recent years [21]. In addition, according to a 2019 survey by the Ministry of Internal Affairs and Communications, 5.1% of those in their 20s and 30s, who use SNSs, have a fixed phone[22]. Therefore, it is easy to be damaged if leaked, but the possibility of being damaged is low in the first place.

⑥ For example, age, date of birth, own photo.
- Age and date of birth are information that cannot be changed, and are information that can be easily identified. However, it is almost impossible to identify a person with only age and date of birth.
- Photographs are information that cannot be changed. It is also information that can easily identify an individual. For example, a photo of a person wearing a uniform may clearly state the name of the school or company. However, before posting on social networking sites, we often consider the account to which we are posting and limit the number of followers who can see the post.

⑦ For example, email address, mobile phone number.
- Both e-mail addresses and cell phone numbers are information that is prone to secondary damage, and both are information whose ownership has increased in recent years with the proliferation of cell phones. However, both are information that can be changed on the same day.

⑧ For example, address, school name / work place, name, family / friend's name
- Addresses, names of schools and companies, and one's own name are information that can be used to identify one's own person, and are vulnerable to secondary damage. However, none of this information is impossible to change.
- It takes a minimum of 3 days from application to approval for a change of address, and a week to 10 days for long-term applications.
- The names of schools and companies cannot be generally changed because of the emotional impact of changing schools or companies, but this information is not impossible to change.
- Although it is necessary to prove that the person is free to change his/her name, it is possible to change the name because there have been cases in the past where people who were slandered on the Internet and suffered from depression or sleep disorders changed their names. It is also possible to change one's name within two weeks to one and a half months after filing a petition to change one's name with the court.
- The names of family members and friends are the same as one's own name.

⑨ For example, medical records. Highest risk.
- Information such as medical records is Sensitive Information with an extremely high degree of privacy, and is vulnerable to secondary damage, so that when it is leaked, it causes significant psychological and economic damage.

In addition, Indicators of Yamada-Model in Table 2 is applied as it is to this paper. "Personal identification" has three values of judgment criteria.
- If it is possible to easily identify an individual with "name" and "address", the value is "2.8291".
- If there is a possibility that an individual can be identified by adding the cost with "name" or "address + phone number", the value is "1.5158".
- If it is difficult to identify other than the above, the value is "1".

Table 4 summarizes the contents of this report. The formulas used in this paper are based on Table 4.

Table 4 Formula Coefficients

Indicators	Comparison	Three-values coefficient		
Changeable information	New-Model	1	1.0129	1.0261
Searchable information	New-Model	1	1.1723	1.3743
Personal identification	New-Model	1	1.5158	2.8291

Applying the values in Table 4 to Table 3 results in Table 5. Each value is the sum of "Changeable information" and "Searchable information". This is the factor for each information type.

Table 5 Coefficients for each information type

		Changeable information		
		Easy (1)	Normal (1.0129)	Difficult (1.0261)
Searchable information	Easy (1.3743)	⑦ Email address, mobile phone number (2.3743)	⑧ Address, school name / work place, name, family / friend's name (2.3482)	⑨Medical records (2.4004)
	Normal (1.1723)	④ Current location / movement history (1.1723)	⑤ Fixed phone number (2.1852)	⑥ Age, date of birth, own photo (2.1984)
	Difficult (1)	①Hobby (1)	②Occupation / Grade (2.0129)	③ Family structure, gender (2.0261)

It is considered that various information leakage information is included in the amount of compensation that comes out in the judgment of the court. And each piece of information has a weight. Also, considering that it is easier to identify the person by leaking multiple types of information instead of simply adding, the following formula is obtained.

- Number of information types: n
- Coefficient for each information type: I_x $(1 \leq x \leq n)$
- Value of any information: V_x $(1 \leq x \leq n)$
- Personal identification: PI
- Amount of compensation for the trial: V_{all}

$$V_x = V_{all} \times I_x / \{(I_1+I_2+I_3+\cdots+I_x+\cdots+I_n) \times PI\}$$

Here, as a premise, even if there are multiple leaked information, if they are of the same type, they are counted as one. Using this newly created formula, the amount of information leaked from the company is calculated.

6 RESULT

The results of a survey on value recognition regarding Personal Data of companies (organizations) and individuals are described.

6.1 Value of Personal Data recognized by companies

Using the coefficients for each information type in Table 5 and the formula in Section 5.3, the amount of money for each information type leaked from the actual compensation for the court reparations was estimated. The following are the main the name of trial and estimation results regarding information leakage (Table 6).

Table 6 Major trials regarding information leakage

Judgment example	Leakage information and amount	Compensation amount
1. 1. NTT phone book case	Name: 12,124 yen Fixed phone number: 11,098 yen Address: 12,124 yen	100,000 yen
2. Uji City, Kyoto Prefecture Resident's card data leakage case	Name: 765 yen Gender: 650 yen Address: 765 yen Date of birth: 705 yen Family composition: 650 yen	10,000 yen
3. Waseda University Ezawa Democratic Seat Lecture List Submission Case	Name: 451 yen Student registration number / university name: 451 yen Address: 451 yen Fixed phone number: 413 yen	5,000 yen
4. Osu City Information Disclosure Ordinance Case	Name: 4,501 yen Address: 4,501 yen Date of birth: 4,145 yen Sensitive information: 4,526 yen	50,000 yen
5. Yahoo! BB customer information leak case	Address: 443 yen Name: 443 yen Mobile phone number: 441 yen E-mail address: 441 yen	5,000 yen
6. Tokyo Beauty Center Questionnaire Response Information Leakage Case	Name: 1,569 yen Address: 1,569 yen Age: 1,445 yen E-mail address: 1,560 yen Occupation: 1,323 yen Mobile phone number: 1,560 yen Sensitive information: 1,578 yen	30,000 yen
7. JAL cabin crew monitoring file incident	Address: 491 yen Name: 491 yen Date of birth: 452 yen Mobile phone number: 488 yen Fixed phone number: 449 yen Gender: 417 yen Sensitive information: 494 yen	220,000 yen
8. Benesse Corporation Customer Information Leakage Case	Name: 176 yen Gender: 150 yen Family structure (scheduled delivery date): 150 yen Parent's name: 176 yen Address: 176 yen E-mail address: 175 yen	3,300 yen

Since ① and ④ in Table 3 did not exist in the trial results in Table 5, the price of each Personal Data was estimated from the judgments of 13 items instead of 15 items. Table 7 shows the average of these. From these results, the one with the lowest amount is the name of a family member / friend, and the one with the highest amount is a fixed telephone number of less than 4000 yen. For example, if the allowable range is less than 1000 yen, the information from occupation / grade to fixed telephone number is highly valuable, so it is better to avoid exposing it to SNSs as much as possible.

Table 7 Value of Personal Data recognized by companies (organizations)

Target Personal Data	Average unit price (yen)
Names of family and friends	176
Family structure	400
sex	405
School name / work place	451
My photo	452
email address	746
mobile number	830
Occupation / grade	1323
Age / date of birth	1678
Sensitive information	2199
Full name	2465
address	2565
Fixed phone number	3987

6.2 Value of Personal Data Recognized by Individuals

We conducted an online questionnaire on the value of personal data for 100 university students in their second to fourth year at Chiba Institute of Technology.

The Price of Personal Information

Suppose your personal information is leaked to a stranger through your contribution. What would you estimate the price of the leaked personal information to be? (If you have never posted before, please use your imagination.) (If you have never posted before, please answer in your imagination.

Interests
Occupation / Grade
Family Structure
Gender
Current location and travel history
Age and date of birth
Photo or video of yourself
Email address
Cell phone number
Fixed phone number
Address
Full name
Name of family member or friend
School name / work place
Sensitive Data

Sensitive data is data with a particularly high degree of privacy regarding physical characteristics, thoughts and beliefs, political or religious views

Figure 2 Contents of the questionnaire

The target audience for the questionnaire is as follows:

- Gender ratio: Male: 62, Female: 35, No response: 3
- Age: 18 to 24 years old (3 people answered that they were in their 20s)
- Department: Chiba Institute of Technology
- When the questionnaire was conducted: November 24th to December 21st

Figure 2 shows the contents of the questionnaire. Amounts are expressed directly in yen. The results are shown in Table 8. The mean and standard deviation are abnormally large due to the influence of some outliers. Although there is no bias, this may be due to the fact that the respondents do not have knowledge of market prices and make judgments based only on their senses without taking market values into account. Therefore, we determined that the mean value does not represent the current situation and verified the results using the median value, which is less susceptible to outliers. The median, like the currency, is a series of "0s," but it is an unadjusted, unbiased value.

Table 8 Values of Personal Data recognized by individuals（Yen）

Target Personal Data	Average value	Standard deviation	Median
Gender	114011.7	998594.3	100
Family Structure	1177083	9993660	1000
Occupation / Grade	9000000000023080000	89548869339593500000	1000
Age and date of birth	1011282249	9999375514	2000
Photo or video of yourself	1790629	10227958	4500
School name / work place	9000000000023080000	89548869339593500000	5000
Email address	2756211	14357866	5000
Names of family and friends	9000000000023080000	89548869339593500000	10000
Cell phone number	22415057	134446428	10000
Sensitive Data	$9.99999999999999\times10^{39}$	$1.00498548997919\times10^{41}$	10000
Full name	$1.00000000000002\times10^{21}$	$9.9498743710662\times10^{21}$	10000
Fixed phone number	4493180	19653382	10000
Address	100106980519	999938392002	50000

The results of the survey are as follows:

- Gender has the lowest amount.
- The amount of money for the address is 50000 yen, which is the highest. In other words, they are afraid that their addresses will be clarified.

6.3 Differences in value perception of Personal Data between companies and individuals

Since the amount of money for companies is based on the opinions of lawyers and experts, and the amount of money for individuals is based on subjective recognition, it is judged that the influence of bias is less for companies and organizations than for individuals. The magnification of individual value recognition is shown based on the company / organization. The results are shown in the table below. (Table 9)

As a result, it was found that the value recognition of individuals is low for gender, occupation and grade, but the value recognition of other items is higher for individuals than for companies / organizations. In other words, it can be seen that the value recognition is higher for individuals than for companies as a whole. In addition, the personal information with the smallest difference in value recognition between individuals and companies / organizations was age / date of birth, and on the contrary, the difference in value recognition of 10 times or more occurred in school names / workplaces and mobile phones. The numbers and addresses, especially the names of family and friends, differed by more than 50 times.

Table 9 Multiplier of individual Personal Data value recognition based on company

No.	Leakage information	Company	Individual	Individual value recognition ratio when the company is set to 1
1	Gender	405	100	0.25
2	Occupation / grade	1323	1000	0.76
3	Age and date of birth	1678	2000	1.19
4	Family Structure	400	1000	2.50
5	Fixed phone number	3987	10000	2.51
6	Full name	2465	10000	4.06
7	Sensitive Data	2199	10000	4.55
8	email address	746	5000	6.89
9	Photo or video of yourself (photo)	452	4500	9.96
10	School name / work place	451	5000	11.08
11	Cell phone number	830	10000	12.05
12	Address	2565	50000	19.49
13	Names of family and friends	176	10000	56.69

7 DISCUSSION

First of all, it can be seen that individuals have a higher recognition of the value of Personal Data than companies. The results can be divided into the following two groups.

A) Information that companies value more than individuals
   - Gender
   - Occupation / grade

B) Information that individuals value higher than companies
   - Age and date of birth
   - Names of family and friends

The direct factor in Group A is that individuals value gender low and companies highly value occupations and grades. In addition, these two pieces of information are highly evaluated by companies because they recognize that they are elements that can be used to attract customers, such as the Recommend function. On the contrary, for individuals, it is "attribute information that cannot be identified by one piece of information", and even if it is leaked, it will not be directly harassed. Therefore, companies recognize that they are more valuable than individuals. It is necessary to recognize that the information in this group is more valuable than the individual thinks.

Next, consider Group B. Leakage of “Photo or video of yourself”, “School name / work place”, “Cell phone number”, “Address” and “Names of family and friends” may be abused by a malicious person. Companies also recognize the important value from the amount of money, but individuals recognize that it is even more important. There are three possible reasons why individual value recognition has become extremely high.

One is fears about the damage that can occur when information is leaked. Second is vague fears that the person actually burned up in the post and was personally identified and learned the fact of slander in the news etc. Third, privacy influences value recognition. Sensitive Information, “Photo or video of yourself”, etc. are information related to privacy. This is less scary than the leak of “School name / work place”, “Cell phone number”, “Address” and “Names of family and friends”, but there is a dislike for the unspecified number of people to know their privacy, so many SNSs It is a factor that boosts individual value recognition of posted information. Compared to corporate value recognition, individual value recognition was excessive. This is different from our assumption, and it was found that individuals have higher value recognition for more information posted on SNSs than companies.

8 CONCLUSION

In this paper, as a result of investigating the value of the information posted on SNSs posts, it was found that the value recognition of individuals is higher than that of companies (organizations) for many information. This is because there is anxiety about damage after information leakage, vague fear of actions that users do not want, and invasion of privacy. In addition, it was found that information related to attracting customers, such as gender, has a higher recognition of the company's value than individuals.

As future issues, since the tolerance is 10% in this questionnaire, we will increase the number of respondents, reduce the tolerance, investigate whether there is any change in the results, and the influence of the difference in value recognition between companies and individuals. At the same time, in order to reduce the number of outliers, we plan to review the questions and content of the questionnaire.

REFERENCES

[1] How to control your Twitter experience, https://help.twitter.com/en/safety-and-security/control-your-twitter-experience

[2] E. Christofides, A. Muise and S. Desmarais, Risky disclosures on Facebook：The effect of having a bad experience on online behavior, Journal of Adolescent Research, Vol.27, pp.714-731(2012).

[3] Brian Stanton, Mary F. Theofanos, Sandra Spickard Prettyman and Susanne Furman, Security Fatigue, IEEE Computer Society, pp.26-32 (2016)

[4] Amended Act on the Protection of Personal Information, https://www.ppc.go.jp/files/pdf/APPI_english.pdf

[5] Difference between “personal information” and “privacy”, https://privacymark.jp/wakaru/kouza/theme1_03.html (Japanese only)

[6] Japanese Industrial Standards, JIS Q 15001:2017 Personal Information Protection Management System – Requirements (Japanese only)

[7] What is personal data?, https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en

[8] Part 1: Special Feature: The Data-driven Economy and Social Transformation, https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h29/html/nc121100.html (Japanese only)

[9] Promoting the Utilization of Personal Data and Ensuring Consumer Confidence, https://www.ppc.go.jp/files/pdf/report_office.pdf (Japanese only)

[10] Report of the Study Group on the Use and Distribution of Personal Data, https://www.soumu.go.jp/main_content/000231713.pdf (Japanese only)

[11] Lee Humphreys, Phillipa Gill, and Balachander Krishnamurthy, How much is too much? Privacy issues on Twitte, Proc. ICA’10, pp.1-30 (2015).

[12] Kenichi Ishii, The “Strong-tied” SNS and “Weak-tied” SNS: A Comparison Regarding the Disclosure of Personal Information and Personal Relationships, Information Processing Society of Japan, Vol.29 No.3, pp.25-36 (2011) (Japanese only).

[13] Hirotsune Sato and Naoya Tabata, Creating an Internet version of the Privacy Dimension Scale, Japan Society of Personality Psychology, Vol.21 No.3, pp.312-315 (2013) (Japanese only).

[14] Shimon Machida, Tomoko Kajiyama, Shigeru Shimada and Isao Echizen, Settings of Access Control by Detecting Sensitive Data Leaks in SNS, Information Processing Society of Japan, Vol.55 No.9 2092–2103 (2014) (Japanese only).

[15] Kenichi Ishii, Uses and Gratifications of Online Communities in Japan, Observatorio, Vol.2 No3, pp.25-37 (2008).

[16] Cheung,Christy,M.K. , Chiu, Pui-Yee and Lee, Matthew K.O, Online social networks: Why do students use facebook?, Computers in Human Behavior, Vol.27 Issue4, pp.1337-1343 (2011).

[17] Nicole B.Ellison , Charles Steinfield and Cliff Lampe, The Benefits of Facebook “Friends”: Social Capital and College Students’ Use of Online Social Network Sites, Journal of Computer-Mediated Communication, Journal of Computer-Mediated Communication, Vol.12 Issue4, pp . 1143-1168 (2006).

[18] Tehila Minkus, Kelvin Liu and Keith W.Ross, Children Seen But Not Heard: When Parents Compromise Children's Online Privacy, Proceedings of the 24th International Conference on World Wide Web, pp.776-786 (2015).

[19] Tomohisa Ishikawa and Kouichi Sakurai, A Study of Compensation in Personal Information Leakage, Computer Security Symposium 2014, pp.1185-1191 (2014) (Japanese only).

[20] Calculating the Cost of a Data Breach in 2018, the Age of AI and the IoT, https://securityintelligence.com/ponemon-cost-of-a-data-breach-2018/

[21] Sasha Romanosky, Lillian Ablon, Andreas Kuehn and Therese Jones, Content analysis of cyber insurance policies: how do carriers price cyber risk?, Journal of Cybersecurity, Vol 5, Issue 1, pp.1-19 (2019).

[22] J. Gregory Sidak and Jeremy O. Skog, Using Conjoint Analysis To Apportion Patent Damages, The Federal Circuit Bar Journal Vol. 25, No. 4, pp.581-620 (2016).

[23] Hong Kwon, Eun-Ju Lee, Tae-Sung Kim and Hyo-Jung Jun, Estimating Compensation for Personal Information Infringement in Korea Using Contingent Valuation Method, Journal of the Society for the Protection of Intelligence, Vol.22 No.2, pp.367-377 (2012) (Korean only).

[24] Investigation Report on Information Security Incidents, https://www.jnsa.org/result/incident/data/2016incident_survey_attachment_ver1.0.pdf

[25] Michihiro Yamada, Hiroaki Kikuchi, Naoki Matsuyama and Koji Inui, Proposal on a Mathematical Model to Estimate Loss of Leakage of Personal Data, Information Processing Society of Japan, Vol.60 No.9, pp.1528-1537 (2019) (Japanese only).

Session 6:
Communication
(Chair: Katsuhiko Kaji)

Organization Training Formulation Method in Project Management - Building a Skills Matrix with Rubric Assessment

Akihiro HAYASHI†, Michi KOMURA, Midori ISHIHARA

Department of Information Design, Shizuoka Institute of Science and Technology, Japan
Osaka Metropolitan University, Japan
ATAS Laboratory, Japan
{pixysbrain, mitikomm347, midorippa2021}@gmail.com

Abstract - Development projects in companies have the critical task of employee development in addition to project goals such as improving QCD. The members assigned to a project are given goals, and they learn by experience through on-the-job training (OJT) to acquire the skills necessary for their respective organizations. Currently, CMMI and PMBOK, which are used as standards for project management, refer to organizational training, but it is not clear how to proceed with OJT in development projects and how to evaluate the results of such training. This study proposes a method to formalize OJT in project management by using rubric evaluation.

Keywords: Project Management, OJT, Rublic Evaluation

1 INTRODUCTION

The organization or projects that develops a system has the critical task of training its employees. It is as important as providing products and services to customers.

Employee training in a company mainly consists of group training and on-the-job training (OJT).When new graduates join a company, there is a training period of several weeks to several months. Training is provided mainly through group education. However, since only textbook content is taught in classroom training, specialized business knowledge is transfered through OJT through real work after being assigned development projects. OJT is defined as an educational method in which employees are given appropriate roles and responsibilities within a project, and are guided by their supervisors and senior staffs in the workplace to acquire knowledge, skills, etc. through actual work.For example, in the case of a software development company, programming languages can be taught to freshmen through group training. Many interactions with long-time customers are undocumented. Such content is passed on through OJT.

However, previous studies have not established best practices in organizational training. The method of training may vary widely depending on the nature of the company's business and organization.In some cases, common content is provided to employees in accordance with training manuals, while in other cases, as in the old apprenticeship system, employees are told to watch and learn from their seniors.

When planning a development project, a deadline is set just in time for the estimated man-hours. There is usually not enough time or cost to meet these deadlines. Even if employees set goals and create training plans within the project, they cannot charge the customer for the time and cost of the training, which is added to the project cost. The plan often falls through.

Even when organizational training is conducted using classroom lectures or on-the-job training, the evaluation methods are often not clear. They may be evaluated based solely on the subjectivity of the observer, or from perspectives that the learner cannot even imagine. Sometimes, personnel evaluations may be mixed.In order to avoid such a situation, it is considered necessary to provide clear indicators and measures before undertaking organizational training, to confirm each other's training goals, and to have a system to evaluate the level of achievement upon completion of the project.

In this study we propose a method that applies the concept of rubric evaluation to organizational training in development projects to make it efficiently and effectively. Rubrics are a system of evaluation used for quality assurance in university education.It uses rubrics consisting of measures of the success of learners' performance and descriptions of the characteristics of each measure.The rubric is used as an indicator to evaluate the achievement of learners by presenting a scale of qualities and abilities required by the organization to learners in advance.

Similarly, the rubric is provided at the start of organizational training to activate communication between instructor and learner by confirming training objectives. Provide timely feedback to avoid one-way teaching and learning, and assess learner's achievement in accordance with the agreed indicators at project completion.

As for previous studies in this field, Ozawa et al[1] pointed out that classes that emphasize learner autonomy are implemented rather than one-way lectures from teachers, such as problem-based learning and learner-constructive classes. They pointed out the importance of the learning environment in learner-constructed classes. Hayashi and Terashima et al[2] clarified the evaluation criteria between learners and evalu-

ators as an issue in measuring educational outcomes.They point out the adverse effects of emphasizing evaluation information from the observer and unilateral grading from the evaluator.Nagata et al[3] report a practical example of incorporating portfolio creation activities and evaluation in a teacher training course, using a class in a different grade level.There have been studies using rubrics for educational activities and learner self-evaluation at universities and other institutions.There is no particular discussion of the application of rubrics to organizational training in system development projects to improve the process.

In this paper, Chapter 2 describes the current status and issues of organizational training in system development projects. In Chapter 3, we explain the concept of rubric evaluation to solve the issues clarified in Chapter 2. Chapter 4 describes future issues.

2 CURRENT SITUATION AND ISSUES OF ORGANIZATIONAL TRAINING ON DEVELOPMENT PROJECTS

2.1 Organizational Training at MMI

CMMI (Capability Maturity Model Integration) is currently the most commonly used process management method in companies, and it describes the basic behaviors that should be encouraged in organizations and projects that develop systems. 22 processes are presented as best practices in five levels, and the maturity levels are graded.One of the processes in CMMI Maturity Level 3 is ”Organization Training”. It describes the training to be conducted in organizations and projects.

Organizational training in CMMI is described as PDCA (Plan, Do, Check, Act) of process management like other processes. By identifying the training needs of the organization and one-year organizational training is planed. Then, t raining is conducted in accordance sith the plan, and the degree of achievement of organizational training is evaluated at progress meetings and project completion meetings. Corrective actions are taken if nessesary when you see a difference between the plan and the actual results.

Organizational training includes group training and on-the-job training.Standardized content with textbooks, such as ITSS (Skill Standard for IT Professionals), is suitable for group training.The company defines the necessary skill sets, the internal training committee prepares training materials, and group training is conducted through classroom lectures by in-house instructors.

On-the-job training is best suited for specialized content that has been passed down from generation to generation in an organization. OJT requires more experience and is not always easy to learn. For example, even if a new employee understands the operation manual, there may be a difference in work speed and product quality between a new employee and a skilled employee. This is called a learning curve, and work efficiency increases as more experience is gained. This is due to the acquisition of tacit knowledge through experience. Newer employees, who still have less experience, lack it. This difference occurs. There is also a ”sense” or ”feeling” that comes with experience.In order to develop skilled employees to the level of skilled employees, the skills and techniques of skilled employees must be transferred to younger employees as well.

However, most skills and technologies are tacit knowledge of skilled employees. It is not easy to transfer tacit knowledge to others.Since it is difficult to transfer tacit knowledge, there is no choice but to have young employees acquire tacit knowledge through experience, just as skilled employees do.Since it is difficult to hand down tacit knowledge to others, young employees, like skilled employees, have no choice but to acquire tacit knowledge through experience.Since it is difficult to hand down tacit knowledge to others, it is necessary to have young employees acquire tacit knowledge through experience in the same way as skilled employees.

CMMI is the most reliable best practice in process improvement. It describes methodologies. It tells you what to do to help your organization acquire the skills it needs, but not how to do it. Even with CMMI in place, there is no established methodology for successful organizational training.

2.2 Requirements for Organizational Training

In this study, we refer to the players in organizational training as ”instructors”, those who teach, such as lecturers, senior employees, skilled workers, and mentors; ”learners”, those who receive training, such as new hires, young employees, inexperienced workers, and mid-career hires; and ”evaluators”, those who judge the improvement of learners’ skills. This section summarizes the requirements for establishing an organizational training process by dividing the organizational training into return on investment, instructors, learners, and evaluators.

2.2.1 Return on Investment(ROI)

Providing organizational training in a company means sacrificing direct hours to make money in the company’s core business and providing training as an indirect operation. From the organization’s point of view, providing organizational training is an investment. Since it is an investment, it is necessary to measure the ROI to ensure that the investment has been worthwhile.

For example, in cases where a company pays the examination fee to take a qualification examination, the investment has been recovered if the employee passes the examination and obtains the qualification. Group education conducted in classrooms and on-the-job training (OJT) within a project are similar investments. There is no clear means to measure the

effectiveness and determine that the investment has been recovered.

The concept of ROI for organizational training differs depending on the position of the players. Even if a learner tries to acquire the expected skills, the evaluator may have a different point of view. Meaningful organizational training cannot be expected unless fair, accurate, and honest evaluations are guaranteed.

2.2.2 Instructors

In university education, active learning and role-playing have been introduced instead of one-way mass-production education. In-house training is not limited to classroom lectures, but also includes a variety of exercises. This is to make sure that the trainees have fully digested the content of the training.

Even in the case of OJT, the assigned roles and responsibilities in the project structure chart are presented to clarify the skills to be acquired by the learner. Evaluate whether the learner can perform the work to earn as a professional in the development project.

Some instructors, however, are more of the apprentice type, where the learner is expected to watch and learn from the work of the instructor and other skilled workers. In such cases, the instructor knows the skill set to be acquired by the learner. However, because it is tacit knowledge, it is difficult to transfer to the learner and does not lead to the expected results. Therefore, the instructor is required to clarify the content of the training in advance and to agree with the learner.

2.2.3 Learners

Organizations provide training to learners in order to help them acquire the skill sets they need. For example, project manager skills, configuration manager skills, advanced C programming skills, meeting minute taking skills, and so on. From the learner's perspective, the purpose of organizational training may be to study in an area of interest, to study for a certification, or for a career path to self-fulfillment.

The skills required by the organization do not always match the skills required by the learner. Providing one-way educational content may not be necessary for learners. Learners do not always overtly promote their own skill sets, leading to a discrepancy between instructor and learner awareness.

To efficiently and effectively conduct organizational training, learners self-assess and disclose their skill sets. Agreement on the skill sets and content to be taught and appropriate feedback during learning are necessary.

2.2.4 Evaluators

Evaluators need to evaluate organizational training one step above instructors and learners; to assess ROI, examine how much the training cost. Ensure that the technicians who complete the training are performing reasonably well on the next project. For instructors, assess whether they are skilled enough to fulfill the roles of instructor, educator, and mentor. If not, a train-the-trainer is needed. For the learner, the results of the organizational training are assessed and registered in a skills matrix to evaluate the effectiveness of the organizational training.

An important aspect of the evaluator's perspective is to distinguish between the role of the project and the evaluation of organizational training. Evaluators evaluate performance against project roles. Project performance and organizational training outcomes do not necessarily coincide. Both instructors and learners prioritize the role of the system development project. And, organizational training evaluations and personnel evaluations should not be mixed.

2.3 Issues to be solved

Leaders, learners, and evaluators are involved in organizational training from different perspectives. To evaluate organizational training, it is necessary to measure that the results are commensurate with the investment.

The issue to be resolved is to establish a methodology for efficient and effective implementation of organizational training.

3 ORGANIZATION TRAINING FORMULATION METHOD USING RUBRIC ASSESSMENT

In Chapter 3, we propose a method of organizational training formallation using rubric assessment to solve the issues presented in Chapter 2.

3.1 What is Rubric Evaluation?

A rubric is a "descriptive form of evaluation criteria that consists of a scale indicating the degree of success of a learner's performance and descriptive words describing the performance characteristics found in each scale"[1].

Rubric assessment, widely used in international university education, increases fairness, objectivity, and planning in the grading of courses. It has been confirmed to be effective for regular evaluation through prior presentation of evaluation criteria and feedback.

3.2 Suggested Rubric for Organizational Training

This study aims to realize efficient and effective organizational training by applying rubric evaluation, which is widely used in universities, to development projects.

		Fully Satisfied	Largely Satisfied	Partially Satisfied	Not Satisfied
Management	Project Management	Has systematic knowledge of project management , self-reliant, can be apointed to PM with confidence.	Can play the role of sub-leader of the project and manage the project in place of the PM in his/her absence.	Understands his/her position within the project and contributes to the team without leadership	Understand only their own work within the project and unaware of the work of other members and dependencies
	Process Management	Familiar with Organization Process (CMMI Level 3) and implement organizational and project processes	Familiar with Project Process (CMMI Level 2) and enable to implement project management by processes	Self-organized and not a Repeatable method of process management.	Process management is ad hoc.
Egineering	Development Skill	Broad development skills of organization's operations and contribute skills to most development projects	Skills contribution in all life cycles defined by organization.	Contribute skills in specific processes of the lifecycle defined by organization	Only possess development skills from previous experienced projects.
	Business Knowledge	can entrust PM in technical fields of no experience.	Contribute to projects while learning new knowledge, in technical fields of no experience	Contribute to projects by replacing and applying similar knowledge/skills in technical field of no experience	Participate in projects of experince only
Commuication	Progress Meeting	Understand master schedule dependencies, proactively manage meetings, and manage risk	Understands Project situation and actively participates in discussions among other members	Prepare reports on their own progress and make necessary reports at meetings.	Participates in meetings, but rarely speaks up and only listens.
	Stakeholder	Ongoing actions to build and maintain good relationships with internal and external stakeholders	Understand what stakeholders are involved and whom to report and consult	Maintains good relationships with project members and participates in the project	Chemistry with the organization or project members

Figure 1: Rublic for Organizational Training

Using the rubric, learners will have a clear guide for learning activities and self-assessment in organizational training. This leads to the discovery and improvement of their own learning issues. By clarifying achievement levels, rubrics can also be used to evaluate management skills and communication skills that are difficult to assess. By using the rubric to reflect on their own learning, learners maintain their motivation for learning and sense of purpose. To make self-evaluation and self-improvement a routine, and to encourage students to learn proactively.

In this study, we propose a rubric for organizational training as shown in Figure 1, targeting organizations that are executing system development. In Figure 1, the subjects of the evaluation indicators on the left vertical axis are the skill items required by the organization. The major items are management, engineering, and communication. Management includes project management and process management, engineering includes development skills and business knowledge, and communication includes progress meetings and stakeholders. The evaluation criteria on the upper horizontal axis describe the evaluation perspective for each of the middle items on the FLPN (Fully, Largely, Partially, Not Satistifed) scale.

The FLPN rating in CMMI is Fully (100-86 %), Largely (85-51 %), Partially (50-16 %), and Not (15-0 %). This is not a numerical value in a quantitative sense, but rather a somewhat sensory evaluation of the degree of skill when a perfect score is set at 100 %. Not only CMMI, but also ISO/IEC 15504 and other standards use a 4-point scale. ISO/IEC 15504 also uses a 4-point scale.

3.3 Organization Training Process Using Rubric

The organizational learning process based on rubric evaluation proposed in this study consists of three phases. Each phase is described in detail in the following sections.

3.3.1 Phase 1

Phase 1 is to present the rubric rating scale in advance, so that the instructor and the learners can clearly define the rubric's points of view. The rubric is then shared with the learner. The rubric shown in Figure 1 is somewhat abstract as an evaluation scale. However, because it is written in an abstract man-

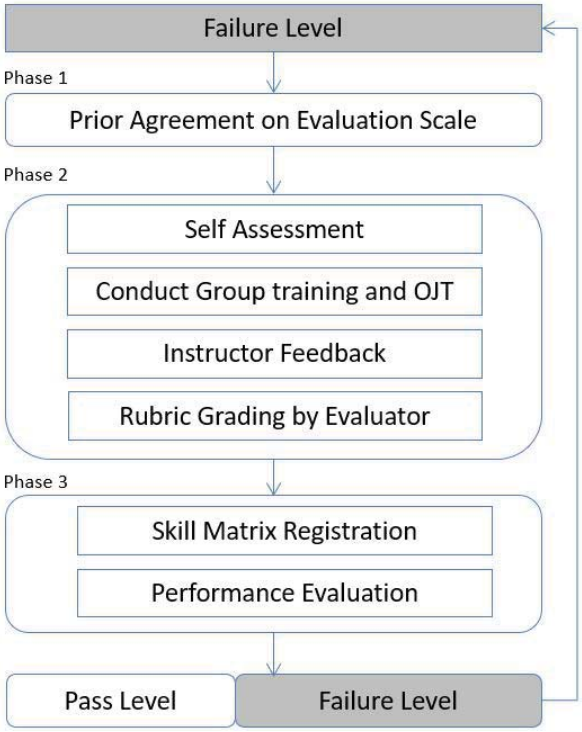


Figure 2: Organization Training Process Using Rubric

ner, it can be applied to the development projects that you are in charge of by interpreting and deciphering the rubric.

3.3.2 Phase 2

Phase 2 describes how to actually use the rubric within the assigned development project.

Self Assessment The learner self-assesses the six medium items as objectively as possible according to the rubric provided, and evaluates them on the FLPN. This allows them to visualize their own rubric.

In the FLPN evaluation, F and L are strengths, and P and N are weaknesses. The self-assessment results show that the P and N are the weak areas of one's skill set, and that by focusing training on these areas, one can acquire a well-balanced skill set.

Conduct Gruop Training and OJT Throughout the appropriate period of time, organizational training is conducted through group training or on-the-job training, according to the organizational training plan. Self-study time is not included in organizational training.

The rubric shown in Figure 1 evaluates the degree of skill acquisition required by the organization using the FLPN, and shows not only the final form of the acquired skill set, but also the intermediate stages of the skill set in the form of Lagely, Partially, and Not Satisfied stages. Learners can objectively grasp the level of proficiency of their own skill set, and the rubric serves as a checklist to guide them to the next step[4].

Instructor Feedback Instructors and learners share the rubrics and their evaluation perspectives in advance and are aware of the purpose of organizational training. However, once the work begins, deadlines and deliverables must be met, and the rubric perspective is disregarded.

Instructors provide feedback at appropriate times, such as at the end of the lifecycle, to let them know what the final evaluation is likely to be if they continue at the current pace.

Rubric Grading by Evaluator After the development project, the evaluators, with the input of the instructor, will give a rubric grade of the learner's proficiency level. Free-text comments are also allowed, and comments are included to help with future improvements.

3.3.3 Phase 3

Phase 3 is the phase in which the rubric grading results are registered in the skills matrix and a pass/fail decision is made for organizational training.

Skill Matrix Registration The FLPNs, which were rubric-graded by the evaluators in the second phase, are used as quantitative values to measure employee growth. The FLPN ratings are graphed as F=4 points, L=3 points, P=2 points, and N=1 point, respectively. The system manages the progress of each employee's skill development to the set goal. Figure 3 below shows an illustration of the system (this is only an illustration, not an actual measurement).

The graph on the left plots the skill development of the engineers every six months. The employees' skills have progressed to the values that they set as their skill improvement targets. The right figure is a radar chart for the middle items of the rubric presented in Figure 1. The rubric indicates the skills that the organization requires of the learner, but not necessarily the skills that the learner wants to acquire. It is also useful to identify the type of engineer.

For example, in an organization that adopts the staff engineer system, all employees who join the company as new recruits are promoted along the same route until they reach the rank of chief.

However, when moving up from the chief position, employees can choose either the management course or the staff engineer course. One is the managerial course, in which the employee progresses from chief to section manager. The other is a course in which the employee receives the same salary as a manager but has no subordinates and is promoted as an engineer.

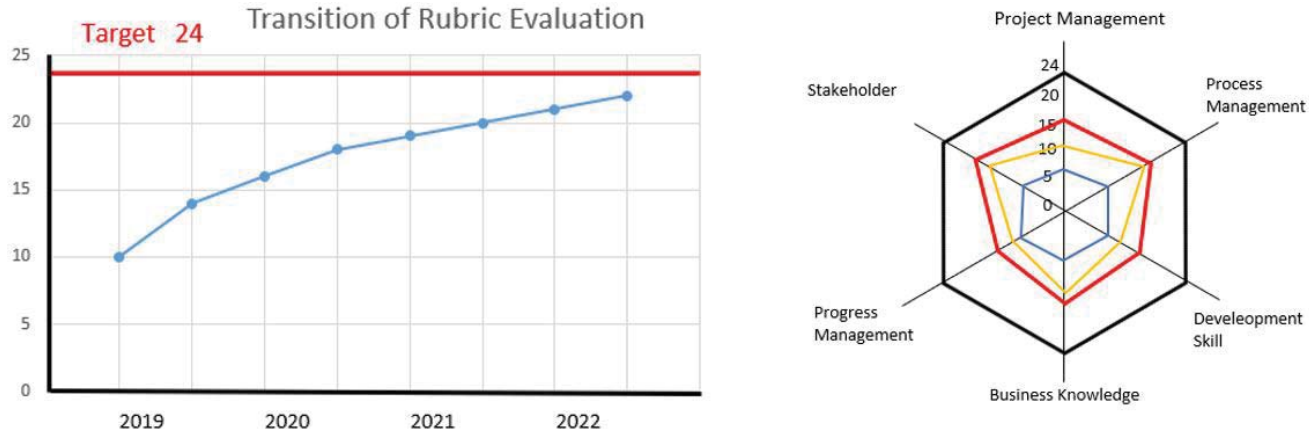


Figure 3: Graph of skill development based on rubric (image)

This is because a good engineer does not necessarily become a good manager.It is unwise to force an engineer who is not suited for a managerial position to become a manager. Even if a good engineer becomes a manager, if he or she becomes too busy with management duties, he or she will not have time to engage in development work and will not be able to take advantage of the high skills that are available in the company. From an organizational point of view, this is equivalent to losing effective resources. Such a type of engineer may choose to progress from staff engineer to senior staff engineer. Adopting the rubric would be effective for this type of job-type decision.

Performance Evaluation Finally, the organization training is a go/no go decision. Organizational training is an investment, and we want to ensure that the skills are developed to the target level in a single organizational training session.

However, as Figure 1 shows, it will take nearly 10 years for a new employee to develop his or her skills to the level of a project manager.It is not possible to create a PM with only one or two organizational training sessions.It is not possible to create a PM with only one or two organizational training sessions. After rubric grading by the evaluator and registration in the skills matrix, performance is evaluated.

4 FUTURE ISSUES

This study points out the issues of organizational training and proposes a formalized method of organizational training using rubric evaluation.

The author works as a university faculty member. He is not in charge of development projects. The proposed method in this study has not been applied to an actual development project. The author will further refine the research content, propose it to companies, and measure its effectiveness by evaluating its application in actual work. This is a future task.

ACKNOWLEDGMENTS

In Conducting This Research, We Referred To The Activities Of ATAS (Arts Therapy Activity Study), a collaborative research project on the spread of various arts therapy activities in local communities. We would like to express our deepest gratitude to ATAS.

REFERENCES

[1] Ozawa S et al, Characteristics and Structures of Learning Environment Design in the Learner-Constructing Course : Analysis of Planning the Joint Camp in Distant Inter-Seminar Using a Bulletin Board System, Japan Society for Educational Technology, PP.143-154,2002

[2] Hayashi T. et al, Possibilities for University Classroom Research Focusing on Evaluation Information from Participating Observers, Nagasaki University Center for Continuing and Lifelong Education, Bulletin 5, , PP.171-182, 2006s

[3] Nagata T. et al, A Practical Study of a Course for Prospective Teachers Involving Portfolio Development in a CSCL Environment That Enhances Various Kinds of Social Interactions, Journal of Japan Society for Educational Technology, PP.218-224, 2002

[4] Shunichi Fukuyama, Procedures for implementing checklists that provide guidance for continuous improvements in software processes, Journal of Information Processing, Vol.42,No.3, PP.529 − 541, 2001

[5] Nishioka K. Rublic, Rublic, Gendai Kyoiku Houhou Jiten, Toshobunka. PP.293, 2004

Proposal for Continuous Improvement of Empowerment Arts Therapy - A Method of Self-Assessment by Applying Rubric -

Michi Komura*, Midori Ishihara**, Akihiro Hayashi***
*Osaka Metropolitan University, Japan
**ATAS Laboratory, Japan
***Shizuoka Institute of Science and Technology, Japan
{mitikomm347, midorippa2021, pixysbrain}@gmail.com

Abstract - It is required to establish a method of objectively grasp the capabilities of practitioners (individuals or groups) of Empowerment Arts Therapy that are based on arts therapy or the therapeutic effects of art, which is conducted in local communities etc., to improve the quality of activities. In this study, we propose a self-assessment method of capability of practitioners by applying rubric evaluation system, which can reveal their strengths and weaknesses and identify areas to improve. This method will enable practitioners to conduct continuous improvement for their capabilities and the quality of activities.

Keywords: Empowerment Arts Therapy, Rubric assessment, Social activity

1 INTRODUCTION

Since 2012, the author has been engaged in a collaborative study on the expansion of various kinds of arts therapy activities in local communities (Arts Therapy Activity Study, hereinafter referred to as “ATAS”).

Arts therapy is a type of psychotherapy that uses artistic techniques such as painting, music, drama, dance, poetry etc. In recent years, a variety of activities based on arts therapy or similar activities to it have been increasing in local communities, not only as psychotherapy or psychotherapy. ATAS distinguishes these types of activities from arts therapy as psychotherapy for therapeutic purposes. ATAS has defined them as Empowerment Arts Therapy (hereinafter “EAT”) and has been working to understand its status quo and issues, and to clarify the characteristics and compositional requirements of its activities.

Empowerment arts therapy aims to support and empower people with relatively higher level of mental health, rather than to treat people with mental illness. It is also conducted in an area that is closer to daily life and encompasses a wider range of people than clinical arts therapy which is based on the relationship between therapists and clients. These various EAT activities are thought to play a certain role in improving people's mental health and Quality of Life (hereafter “QOL”), and in supporting and empowering people with issues in the community.

On the other hand, many of these EAT activities face difficulties in securing funding and human resources, and in sustaining their activities. Social recognition of the activities and their practitioners is vague, and there are no established criteria or guidelines for judging the skills of the practitioners and the content of their services.

In this study, the authors propose a rubric-based self-assessment method for continuous improvement of EAT.

Chapter 2 outlines previous studies on EAT and the issues that need to be resolved. Chapter 3 describes the significance and expected effects of rubric-based self-assessment. Chapter 4 presents future issues and research prospects

2 BACKGROUND AND ISSUES IN EAT STUDIES

2.1 Previous Studies on EAT

In 2008, the Konan Institute of Human Sciences (KIHS) at Konan University launched a joint research project, “Interdisciplinary Research for Establishing a Common Platform for Study of Arts and Arts Therapy.” As a part of the project, a questionnaire and interview survey of arts therapists were conducted. As a result, various aspects of arts therapy in Japan and its social significance were clarified. It was confirmed that arts therapy activities exist not only as psychotherapy based on a therapeutic approach, but also as empowerment, mainly for the purpose of maintaining and improving health and QOL, and that they are functioning effectively in local communities. At the same time, the issues related to training of the practitioners of these activities emerged [1].

Based on the results of this research by KIHS, a joint research project targeting various types of arts therapy activities and their practitioners in local communities was newly launched in 2012, which conducted a nationwide questionnaire survey, interview survey (semi-structured interviews), and on-site observation of activities. As a result, the status quo and characteristics of arts therapy in the communities were identified as well as the issues that need to be addressed for its development. Furthermore, arts therapy mainly intending support rooted in daily life was defined as EAT, and the requirements for EAT activities and practitioners were developed in 2017 [1].

The surveys above revealed that the main issues related to EAT are sustainability (management structure) and quality assurance (quality of services, skills of the staff). 65% of the respondents answered they earned "one million yen or less" through arts therapy activities and spend for arts therapy activities for “10 hours or less” per week [2]. Considering this reality, financial independence through EAT activities alone is not feasible in the short term. Therefore, it is

realistic to put financial independence on hold and explore ways to increase the sustainability of the activities.

The internal factors of the practitioners, such as lack of skills and lack of awareness, are issues of higher priority, as they are related to the interests of the users, safety assurance, social responsibility, and above all, the credibility of EAT. Many practitioners are engaged in EAT activities based on their good intentions and initiative. However, there is also a risk of self-righteousness. The practitioners must avoid being unaware of possible risks or making mistakes that could be detrimental to the users because EAT is an activity committed to human psychology. Recognizing possible risks and taking appropriate measures to prevent them are essential to improve the quality of EAT activities.

In ATAS, as a response to the identified issues on EAT, the EAT composition requirements were developed and published as criteria for self-assessment [1]. The reason for adopting the self-evaluation method in ATAS is that, as described below in the Chapter 3.1, since the fields, purposes, methods, scale, and aspects of EAT activities are diverse, establishing a uniform and objective evaluation method would require a large-scale and long-term research and study. On the other hand, it was an urgent issue to prevent errors in the assessment in the implementation of EAT activities. Considering the current situation of EAT, a self- evaluation method that allows the practitioners themselves to work on proactively was adopted.

2.2 Issue to be Solved

The self-evaluation method proposed in ATAS was that EAT practitioners score the degree of achievement of each component, depicts it graphically on a radar chart, grasps his/her own characteristics and position by the result, and discloses it on the Internet, etc. However, it has not been formed as a practical tool for practitioners to use, and it has not yet been implemented. The development of a practical evaluation tool that assesses the capability of practitioners and contributes to the improvement of their skills and the quality of their services has been the issue to be solved.

Therefore, the authors have scrutinized and revised the evaluation criteria and methods by ATAS and proposed the Assessment Sheet for EAT Activities (hereinafter “Assessment Sheet”) [3]. The sheet is based on the component requirements/evaluation criteria clarified through ATAS evaluation. It includes a checklist of 25 items in 8 categories (App. i) and supposed to be utilized as a self-evaluation tool by EAT practitioners. The practitioners are asked to quantify each check item on a 5-point scale, calculate the average value for each category, and enter the average value on a radar chart (App. ii). This allows the practitioners to objectively view their activities and themselves, and to visualize policy, basic stance, characteristics, tendencies, strengths and weaknesses of their own activities.

However, although this evaluation sheet clarifies the level of achievement for each item, it is difficult for the practitioners (i.e., evaluators) to grasp what specific improvements should be made for items with low scores, and what status to aim for as the next step. It is necessary to establish a method to improve the skills of practitioners and the quality of services by having practitioners conduct regular self-assessment and clearly identify areas in need of improvement, and to provide concrete directions for improvement that should be aimed for.

3 METHOD FOR CONTINUOUS IMPROVEMENT OF EAT

3.1 What is EAT?

EAT is an academic descriptive concept or category that comprehensively describes various types of arts therapies in the community, as opposed to arts therapy which is based on clinical and therapeutic approaches. It is not a term intended to be used as a name for individual and specific activities or a title of activists, so EAT practitioners do not call their activities “Empowerment Arts Therapy” or call themselves “Empowerment Arts Therapists” in their actual activities.

The concept of empowerment in social contexts has changed over time, beginning with its use in the civil rights and feminist movements in the U.S. in the 1960s and 1970s, gradually expanding its meaning and applied fields. In the late 1990s, it expanded into the fields such as medicine, health care, welfare, and education. Today, it targets people with difficulties and difficulties in their lives in general, and by working on their inherent and latent resources and abilities, it has come to refer to empowering them to become aware of their resources and abilities, and to accompany them in thinking, choosing, and asserting themselves independently [4].

The introduction of the empowerment concept makes it possible to view arts therapy on two axes: “therapy (or treatment)” and “empowerment”. In ATAS, arts therapy based on clinical approach is distinguished as Psychopathological Arts Therapy (hereinafter “PAT”) from the concept of EAT. The distinction between EAT and PAT makes it possible to organize the relationship between the two and to examine various arts therapy activities in terms of both EAT and PAT elements.

The following facts about EAT were obtained from the study so far [1].

(1) EAT is not a therapy in the narrow sense which means treatment of patients with mental disorders but is intended for the empowerment of people with relatively higher level of mental health. The main targets of EAT are, for examples, those who seek to reduce stress at work or in daily life, those who seek to develop their skills or improve their QOL, and those who have some mental problems but are not at the stage where treatment is necessary.

(2) The fields of EAT activities are diverse, including welfare and education, etc. ATAS has classified arts therapy into five categories, each organized in relation to psychopathological and empowering approaches as follows (Fig.1).

(3) Many EAT activities emerge from within the practice and are developed spontaneously and autonomously. While referring to and applying theories and techniques of psychotherapy as necessary, the practitioners are not bound by existing systems and

methods but add their own unique or original elements to their own practice in response to the needs of the field.

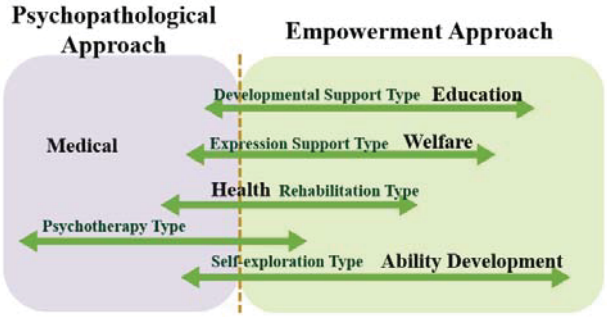


Fig. 1 Areas and Positions of Arts Therapy in Japan

(4) There are few opportunities to receive appropriate remuneration for their activities, and the availability of management resources such as funds and human resources is unstable. This fact makes the sustainability of their activities an issue.

(5) Since there are no institutional standards of behavior and evaluation that define professional requirements, there could be some cases of unsafe or self-righteous behavior by practitioners. This poses the risk of causing unintended psychological disadvantages to users of services or clients.

3.2 Construction of Rubric

As described in Section 2.2, the purpose of this study is to identify more clearly areas for improvement in EAT activities and to develop index to show specific directions for improvement. In this study, we propose a rubric that will serve as a guideline to indicate the direction of improvement to be aimed for.

A rubric is a document that specifies a task description (assignment), a scale of some sort (levels of achievement, possibly in the form of grades), the dimension of the assignment (a breakdown of the skills/knowledge involved in the assignment), and description of what constitutes each level of performance (specific feedback) [5] as the basic elements that indicate learning achievement. In educational settings, it is used by teachers to provide accurate and prompt feedback to students, and by teachers and students to share learning goals and evaluation index. In addition to teachers' evaluation, self-assessment by learners is expected to enhance learners' autonomy, clarify learning goals and contents, and enable them to proactively grasp the results and challenges of their studies [6].

These features of rubrics include the potential to contribute to continuous self-improvement in the field of social actions or enterprises as well as grading in school. However there has not been discussion on the possibility of improvement through utilizing rubrics in these fields in Japan. A rubric is also expected to be applied to self-assessment for EAT practitioners with diverse backgrounds, enabling them to reflect on their skills and activities, objectively grasp their strengths and weaknesses, and contribute to the improvement of their capability.

The purpose of using a rubric for EAT activities is self-assessment by the activity leader and continuous improvement based on this self-assessment.

The advantages of rubric assessment in EAT activities are as follows

1. What is required for each item is clarified.
2. What to specifically do is clarified to improve items that are marked on low levels.
3. Practitioners can share information on what to keep in mind when conducting EAT activities.

This will enable practitioners to identify their own strengths and weaknesses, develop and actively promote the strengths, and formulate plans for improving the weaknesses. By doing this on a regular basis, it is possible to continuously improve and raise the level of activities. It is also possible to form a certain common understanding of EAT activities, which are diverse and difficult to standardize.

The proposed EAT rubric assesses the seven aspects of "Knowledge," "Skill," "Psychological Safety," "Self-exploration / Self-understanding," "Cooperation / Networking," "Sustainability," and "Philosophy / Mission" on a four-point scale (Fig. 2).

The rating scale is "Level 1 (Not Satisfied)", "Level 2 (Partially Satisfied)", "Level 3 (Largely Satisfied)", and "Level 4 (Fully Satisfied)".

Levels 1 and 2 can be identified as areas in need of improvement. Level 1 is what should have been achieved, and items checked here indicate that there is a great deal of room and need for improvement. Level 2 is in the process of improvement. Level 3 is a level that has been generally met, and the practitioners are expected to make concrete efforts to achieve this level first. Level 4 is the highest level that can be desired and is set as an ideal best practice and serves as a goal or guideline for each item.

3.3 Process of Improvement

In this study, the authors propose a method for continuous improvement of EAT using the rubric indicated in 3.2. The method consists of self-assessment using the rubric, identification of areas in need of improvement, formulation of improvement plans, implementation of improvement plans, and self-inspection using the checklist (Fig.3)

3.3.1. Self-Assessment by Rubric

EAT practitioners (individual or by group) periodically self-assess, using a rubric, which rating scale each of their activities fits into.

3.3.2. Identification of Areas in Need of Improvement

Based on the results, the practitioners identify their own Strengths and Weaknesses, and then identifies the areas in need of improvement. Items rated 1 or 2 on the rubric scale are areas in need of improvement. When there are multiple

		Level 4 Fully Satisfied	Level 3 Largely Satisfied	Level 2 Partially Satisfied	Level 1 Not Satisfied
1	Knowledge	Have sufficiet knowledge of arts therapy required in your specialized and related activity areas. Have good comprehension throughout artstherapy and be able to explain it for others to understand.	Have sufficient knowledge of arts therapy required in your specialized activity area. Able to explain it for others to explain.	Have basic level of knowledge of arts therapy.	Knowledge is limited or fragmental. Have not studied arts therapy systematically.
2	Skill	Able to provide safe and effective EAT with originally devised methods as well as methods mastered. Able to customize the methods and process flexibly according to a client's condition.	Able to provide safe and efficient EAT with method mastered.	Able to provide EAT based on the methods mastered.	Have very few or a few experiences of EAT practice.
3	Psychological Safety	Psychological safety in the activities is secured. Unexpected trouble is responded swiftly and propoerly. Able to provide proper alternative solutions such as to introduce the client to other relevant agencies, specialists in case that the matter is beyond your skill.	Aware of the need and importance to secure psychological safety for clients in the activities and necessary measures are taken.	Ready to possible troubes to some extent. Able to see if the matter is beyond your skill at least and take some measures.	No measures for sychological safety is taken in the activities. Do not think of any troubles out of your control that will possibly happen. Have no way to deal with troubles.
4	Self-exploration / Self-understanding	Have sufficient and profound understanding of your deep inside and essential propensity. Your own psychological issues have been overcome. Periodically take training analysis or supervision.	Have sufficient self-understanding of your inside and essential propensity. Your own psychological issues are resolved.	Self-insight about him/herself has been done to some extent.. Aware of his/her psychological issues and working on to find ways to solve them.	Never had any self-insight about yourself.. Your own psychological issues are untouched or unrecognized.
5	Cooperation / Networking	Have reliable connections to contact professional people in the same field as you and other related fields. Belong to academic societies and regularly participate or make presentations in conferences.	Have reliable personal connections or fellows. Gathering information constantly.	Trying to have a reliable personal connections or fellows. Able to get information when you need.	Have no personal connections or fellows to ask or consult in case of trouble. Have no way for gathering information about activities.
6	Sustainability	Working sustainably and constantly, as well as have a possibility of develop of activities.	Working sustainably regardless of whether paid or not.	Working on or making efforts to raise the sustainably of your activities.	Neither think of the sustainability of activities nor the managerial side of your activities for raising sustainability.
7	Philosophy / Mission	The philosophy/mission is firmly established and clearly verbalized. Able to communicate the purpose and ideals of your activities to others.	The philosophy/mission is established and able to clearly explain it to others clearly.	Have the philosophy/mission and verbalized to some extent.	Basic philosophy/mission is not established. Or have never thought of them.

Fig. 2 EAT Rubric

items requiring improvement, they are prioritized and addressed for improvement. Basically, items with a rating of Level 1 are judged to be more in need of improvement than those with a rating of Level 2, but the priority of improvement need not necessarily be judged mechanically. It is advisable to make a proactive decision on what to improve first in the light of one's own activity policy and the points that are important.

3.3.3. Development of Improvement Plan

For the items identified as areas requiring improvement, a list of points to be improved is made, and an improvement plan is formulated by analyzing and discussing what points are insufficient and how they can be improved, determining the steps to be taken by setting priorities, deciding who will be in charge, the period of implementation, and so on. The improvement plan should be as specific as possible. However, it is essential to make a plan that is feasible according to the actual situation of the activities, because it would be a complete reversal if the burden of improvement activities becomes so heavy that it interferes with the primary activities.

3.3.4. Implementation of Improvement Plan

The practitioners engage in improvement activities according to the developed plan. The rubric is not only a self-assessment tool, but also a guidance of activities, presenting basic behaviors that the practitioners should encourage. Following the rubric leads the practitioners to a tireless and steady accumulation of the required activities.

3.3.5. Self-Inspection by Checklist

After carrying out the planned improvement activities, self-evaluation using the checklist on the Assessment Sheet (App. i) is performed and entered in the radar chart (App. ii). By following the changes in the shapes of the radar chart, the process of continuous improvement of EAT activities can be visualized. If the chart shapes become more balanced, it can be confirmed that the improvement is well underway.

Finally, when the radar chart reaches the highest level, the practitioners are supposed to have achieved the level this rubric proposed and can be considered to have fulfilled the required level of EAT activities.

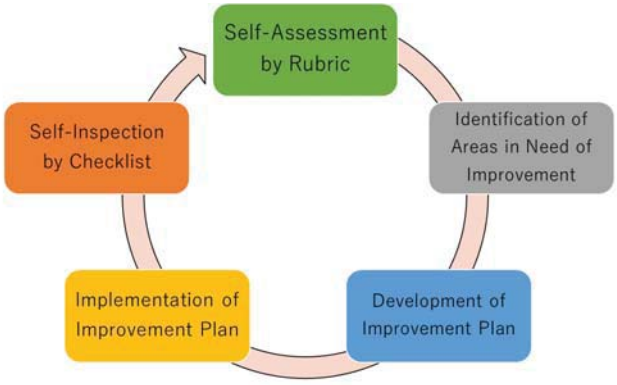


Fig. 3 Process of Improvement

4 FUTURE ISSUES AND RESEARCH PROSPECTS - TOWARD VERIFICATION

In this study, a method for continuous improvement of EAT activities using a rubric was proposed. In the future, we would like to verify the effectiveness of this method and identify areas for improvement by having as many EAT practitioners as possible utilize this method. We plan to collect and analyze as much feedback as possible, by contacting the informants who cooperated ATAS project in the past, developing new informants and obtaining research grants, etc. These will be the subjects of future work, along with the discussion on specific methods for verification surveys.

As Terashima and Hayashi (2006) stated, problems associated with improvement activities based on self-assessment include the possibility that the practitioners are unfamiliar with rubric evaluation and find self-assessment difficult, and that it is easier to find problems than good points, which in turn may lead to a loss of confidence in their own activities[6]. These points need to be given sufficient consideration in the verification surveys to be conducted in the future.

Both rubric and checklist proposed will continuously be revised as necessary. The objective of constructing the EAT rubric is to empower the practitioners to improve their capability through the assessment. The goal is to establish an evaluation index to improve the quality, sustainability, credibility, and social recognition of EAT which are difficult to evaluate uniformly due to their wide variety, while drawing out their good qualities such as diversity and creative potential. We will continue to work on the development and improvement of the tools for this purpose, as well as dissemination and raising public awareness.

REFERENCES

[1] Kaneko, H. & Ishihara, M., *Study on Empowerment Art Therapy Today* [*Empowerment gata art therapy kenkyu no genzai*], The Crisis of the Mind and Clinical Knowledge, Bulletin of Konan Institute of Human Sciences, Vol.20, pp.21-41, Mar.2019 (in Japanese).

[2] Ishihara, M. & Kaneko, H., *Art Therapy Activities in the Communities as Empowerment - Its Endogenous and Autonomous Nature Seen from the Nationwide Survey* [*Empowerment toshiteno shisei no art therapy katsudo – Zenkoku jittai chousa kara mieru sono naihatsusei to jiritsusei*], The Crisis of the Mind and Clinical Knowledge, Bulletin of Konan Institute of Human Sciences, Vol.16, pp.105-130 Feb.2015 (in Japanese).

[3] Komura, M. & Hayashi, A., *A Method of Assessment for Empowerment Arts Therapy*, IEICE Technical Report - Shingakugiho, Vol.121, No.372, SWIM2021-31, pp.1-7, Feb.2022 (in Japanese).

[4] Kaneko, H. & Ishihara, M., *A View of "Art Therapy for Empowerment"*, Bulletin of Kobe University of Welfare, Vol.19, No.1, pp.1-22, Dec.2018 (in Japanese).

[5] Stevens, D.D. & Levi, A.J., *Introduction to Rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback, and Promote Student Learning*, Second Edition, Sterling, Virginia, Stylus Publishing, LLC, Chapter 1, 2013.

[6] Terashima, K. & Hayashi, T., *Development of Problem-Based Learning to Promote Self-Evaluation Using Rubrics*, Kyoto University Studies in Higher Education, Vol.12, Dec.2006 (in Japanese).

Appendix: Assessment Sheet for Empowerment Arts Therapy Activity

Empowerment Arts Therapy（EAT）Assessment Sheet

Ⅰ SKILLS OF PRACTITIONERS

1 Knowledge / Skills

		Marking by yourself	Average
①	Understand and explain the therapeutic effects and functions of art or (artistic) expression.	1・2・3・4・5	
②	Possess sufficient knowledge and skills in arts therapy and art to appropriately conduct activities according to the purposes and aims.	1・2・3・4・5	
③	Knowledgeable about techniques, characteristics, history, etc. of arts therapy outside of your own area of expertise, as well as your own expertise.	1・2・3・4・5	
④	Possess knowledge and skills in psychotherapy, psychotherapy and psychological counseling at the level required in your practical activities.	1・2・3・4・5	

2 Self-exploration / Self-understanding

⑤	Working to improve knowledge and skills required for its activities in your specialized and/or related areas of expertise.	1・2・3・4・5	
⑥	Conduct self-assessment.	1・2・3・4・5	
⑦	Have understanding of the nature and popensities of your own, and your psychological issues are resolved.	1・2・3・4・5	

Ⅱ PRACTICE

3 Practice of EAT

⑧	The purpose, aim, and target (client) of your activity are clear.	1・2・3・4・5	
⑨	Able to accurately set objectives and goals according to the client's situation and needs, and to organize and proceed with work and sessions.	1・2・3・4・5	
⑩	Appropriate psychological distance is maintained with the client, and a stable and trusting relationship is established.	1・2・3・4・5	

4 Rapport with Clients

⑪	Working to understand the clients from multiple perspectives.	1・2・3・4・5	
⑫	Working to obtain information about the client's social environment, and related institutions, organizations, services, etc. needed to support clients.	1・2・3・4・5	
⑬	Questionnaires, interviews, etc. are conducted with clients regarding the effectiveness of EAT.	1・2・3・4・5	

5 Psychological Safety

⑭	Maintain confidentiality regarding the privacy of clients.	1・2・3・4・5	
⑮	Make sure that participants should maintain confidentiality regarding the content and privacy of what they say during group work and sessions.	1・2・3・4・5	
⑯	The necessary individual follow-up is provided according to the psychological situation of the client.	1・2・3・4・5	

6 Professional Network

⑰	Have a supervisor or mentor to consult as needed.	1・2・3・4・5	
⑱	Have information and networks to connect clients to the appropriate experts or professional organizations, depending on the client's condition and the situation of the case.	1・2・3・4・5	

Ⅲ MANAGEMENT STYLE

7 Management

⑲	Engaged in activities regularly and continuously. Or working toward this goal.	1・2・3・4・5	
⑳	Receiving compensation / remuneration commensurate with the activity.	1・2・3・4・5	
㉑	Expenses related to activities are covered by compensation for activities and/or grants, etc., and do not rely on out-of-pocket expenses.	1・2・3・4・5	
㉒	Activity and income/expenditure plans are drawn.	1・2・3・4・5	

Ⅳ VALUE

8 Philosophy / Mission

㉓	The philosophy, mission and the ideals of the activitiy are clearly defined, and able to explain them to others.	1・2・3・4・5	
㉔	The characteristics of your EAT activities are clear and able to explain them to others.	1・2・3・4・5	
㉕	Maintain an attitude to pursue the essence of expression.	1・2・3・4・5	

i. Checklist

ii. Radar Chart

A Study of Progress Management Method using TA-Bot for a Software Development Class

Kocichi Ono*, Ryozo Kiyohara‡, and Yoshiaki Terashima*

* Graduate School of Science and Engineering, Soka University, Japan
‡Faculty of Informatics, Kanagawa Institute of Technology, Japan
e21m5311@soka-u.jp

Abstract - In this study, we discuss a method that uses TA-Bot to manage student progress in software development classes. TA-Bot calculates the severity of developmental delay to monitor student progress and provides technical advice based on its experience of the class to resolve the delay. Moreover, experimentally determining the optimal timing of technical advice and the technical advice as the system threshold for each student requires student participation and repeated experiment. However, owing to the difficulty in creating the intended development situation and forming teams of students, the experiment is not reproducible. Therefore, we repeat experiment using simulation and adjust the system thresholods for each student. For this purpose, we define the student parameters required for faithfully reproducing the development situation of students to intentionally design different situations for each student.

Keywords: Software Development, progress management, TA-Bot

1 Introduction

In this study, the Teaching Assistant Bot (TA-Bot) is used to manage student progress in a software development class. Conventionally, a TA provides technical advice depending on the developmental situation by monitoring student progress. This process is automated using TA-Bot. The class considered in this study is conducted online, where students and teacher share information using social media. TA gets students' progress report and gives technical advice for students experiencing developmental delay. To perform these tasks effectively, the TA-Bot must be provided with definitions of optimal advice and timing. Thus, the developmental delay is quantified into levels and the TA-Bot is provided with technical advice from a database. Moreover, student participation is required to properly evaluate the method. However, the intended development situation cannot be created owing to the disparity in the programing skill levels of the students. Thus, the experiment is not reproducible. To overcome this problem, the experiment is performed on a simulation of the development situation, which can be repeated to confirm each student the effectiveness of the technical advice and determine the optimal timing each student for it.

This paper is organized as follows: Section2 outlines the background of the study. Section3 explains software development class and our proposed method for this study. Section4 explains experiment, parameters reproducing students' development situation, and analyzing the evaluation result. Finally, Section5 concludes the manuscript.

2 Background

Recently, online classes are widely disseminated. Students participation can take the class asynchronous each place. But, they are difficult to communicate with students participation, unlike face-to-face ones. Thus, monitoring state of students' learning is a burden on a TA because it is difficult for the TA to monitor them. In addition, the more students are, the greater a burden on TA.

Education Network for Practical Information Technologies (enPiT) [**?**] is an education project to develop the advanced information technology human resources based on practical education in cooperation between universities and industry promoted by Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan. [**?**] enPiT has four fields; AiBiC, Security, Emb and BizSysD, and software development is increasingly important. Now, enPiT is implemented at Japanese universities, we presume that it will be implemnted at high school and junior high school.

Based on these two things, the study of software development was conducted so far.[**?**][**?**][**?**] So, we focus on TA monitoring students' progress management for a software development class.

3 Software Development Class and Proposed Method

3.1 Software Development Class

The target of this study is a software development class that is conducted online. Figure1 illustrates the structure of the class comprising approximately 100 students. Each student created their developmental plan individually before the commencement of the class and submitted it to the TA. To help the TA evaluate their development progress, the students specify their progress rate and development phase (design, coding, or testing). Finally, each student develops the software asynchronously according to their convenience.

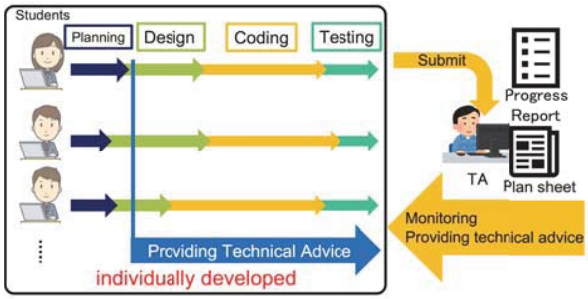


Figure 1: Software Development Class

3.2 Management Problem

Based on the description above, we concluded that the burden on a TA is large. It has to simultaneously monitor the progress of several students, which differ significantly corresponding to their programing skills. In addition, confirming the progress of all the students is difficult. Consequently, the TA cannot provide appropriately timed technical advice and the developmental delay worsens.

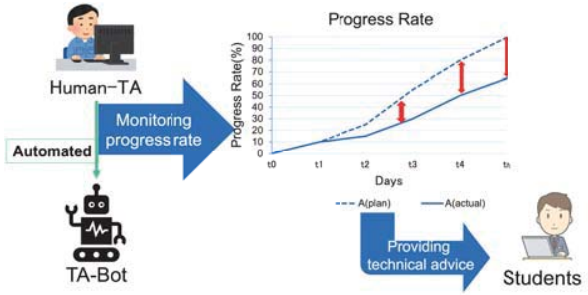


Figure 2: Problem monitoring students' progress

3.3 Progress Management Method

We propose a progress management method using TA-Bot to resolve the aforementioned problems faced by the TA. We automate the processes of monitoring student progress and providing technical advice, thereby lightening the workload of the TA. Accordingly, we define two components for TA-Bot.

1. Monitoring student progress

2. Providing appropriate technical advice to each student

TA-Bot provides technical advice to students experiencing developmental delay. For this purpose, TA-Bot needs to be aware of the difference between the planned and actual progress. However, TA-Bot cannot measure the severity of the developmental delay from this difference. Therefore, it can neither provide appropriate technical advice nor eliminate developmental delay.

Moreover, although the TA-Bot selects technical advice from a database, it does not understand how the advice affects students. In other words, TA-Bot cannot select appropriate technical advice. Therefore, we need to define indicators that qualify technical advice as appropriate. Furthermore, to provide students with the latest technical advice, we need to formulate updated advice for the database.

3.3.1 System Configuration

Figure3 shows the configuration of the proposed system. The social media for communication between students and the instructors that we use is Slack[**?**]. TA-Bot has a plan for each student and acquires technical advice database. The TA, and teacher use Slack to share information with the students, who submit weekly progress reports and specify their progress rate, which indicates developmental progress, and developmental situation in the reports. TA-Bot has two system thresholds for each student; TA-Bot stores the provided information in the database, calculates the severity of developmental delay by comparing the progress rates in the progress report and the plan, and finally selects technical advice based on the severity. This type of advice is called Static Advice. Furthermore, we store technical advice provided in past classes and randomly add new technical advice. TA-Bot identifies keywords related to student development by analyzing the comments on the social media, and searches for the keyword on the web. Finally, TA-Bot stores new technical advice that is acquired from the web in the database. This type of advice is called is Dynamic Advice.

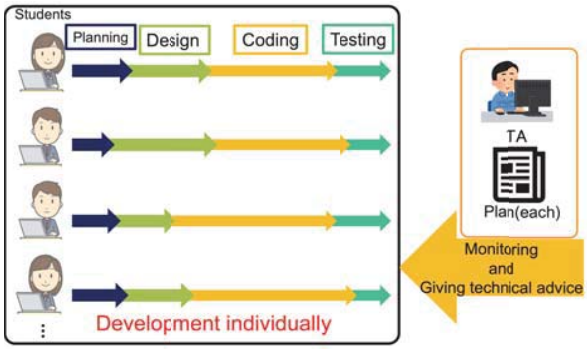


Figure 3: System Configuration

3.3.2 Severity

Severity quantifies developmental delay into separate levels. In other words, the difference between the progress rate in the plan and that in the progress report reveals developmental delay, but not its severity. Figure4 shows how the developmental delay and its severity is quantified. to ensure that the TA-Bot provides more effective technical advice. The severity is calculated by multiplying the difference between the progress rates in the plan and progress report by the number of elapsed days. Generally,

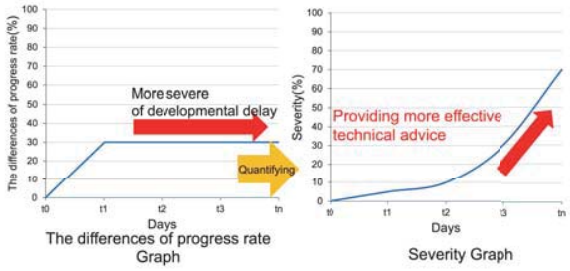


Figure 4: Quantifying the severity of developmental delay

the severity is higher closer to the development completion date. In this context, Tnp is the planned progress rate. Tna is the progress rate in the progress report. E is the number of days elapsed. D is total number of days. The formula to calculate severity is given by (1).

$$\frac{Tnp - Tna}{Tnp} \times \frac{E}{D} \tag{1}$$

3.3.3 Advice

As explained earlier, two types of technical advice, i.e., Static and Dynamic Advice, exist based on the source from which the TA-Bot acquires technical advice.

Static Advice is the technical advice acquired from the database. Figure5 shows how static technical advice is selected. TA-Bot selects the technical advice depending on the calculated severity of developmental delay. For this purpose, we define thresholds to help the TA-Bot in deciding what type of technical advice to provide by splitting the severity into three levels, i.e., Level1, Level2 and Level3. By adjusting the threshold for each student, the optimal timing for providing technical advice can be adjusted. Table1 shows the three types of technical advice, i.e., providing keywords related to software development, directing students to a website that contains a relevant sample of software, and providing a programming syntax. For Level1 severity, TA-Bot provides a keyword that is prepared from its experience of the class. For Level2 severity, TA-Bot directs the student to a website that is finds by scraping based on the keyword. Finally, for Level3 severity, TA-Bot provides the programming syntax to facilitate comprehension and encourages the students to develop the software further. In addition, the database of the TA-Bot is continually updated by the TA, which ensures that the students receive optimal technical advice.

Table 1: Technical Advice

Severity Level	Technical Advice
1	Keyword
2	Web Information
3	Syntax

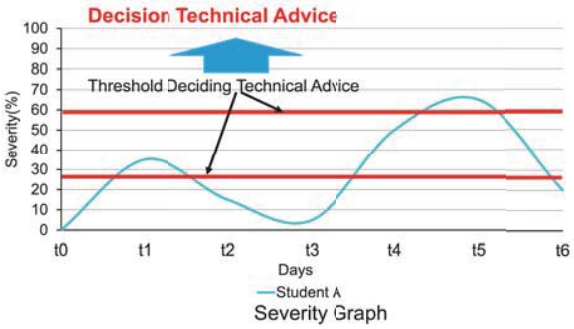


Figure 5: Static Advice

Dynamic Advice is the technical advice acquired from websites that contain information on software development. TA-Bot analyzes student comments and identifies keywords related to the specific problem they are facing in the process of software development. Figure6 shows how to dynamic technical advice is selected. TA-Bot divides each word in the comments using morphological analysis, identifies keywords related to software development, and finds the relevant website by scraping based on the keyword. Finally, it stores the new technical advice in the database and makes it available to the class the next time advice is required. Thus, problems faced by students in software development can be flexibly solved using Dynamic Advice.

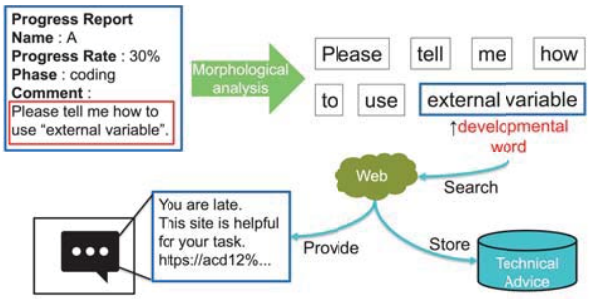


Figure 6: Dynamic Advice

However, the effectiveness of the proposed TA-Bot cannot be easily evaluated because of two reasons: the scale of the required experiment is large and ensuring student participation is difficult. Thus, data can be collected only from a few students, which is insufficient for evaluation. Moreover, the intended development situation is not available for evaluation. Therefore, the experiment is not reproducible.

4 Simulation

To optimize two System Thresholds for each student; Diciding Technical Advice Threshold determing providing the technical advice timing and Selecting Technical Advice deciding which technical advice providing to students, we need to repeat experiment. Accordingly, we ask the students participating in the ex-

periment to solve programming problems, and TA-Bot provides technical advice to each student depending on their developmental situation. Using this process, we repeat the experiment and, thereby, adjust the timing and the effectiveness of the technical advice.

However, because the real experiment is not reproducible, the evaluation is performed using simulation, for which we intentionally create developmental situations for each student. Consequently, we can repeat the evaluation for cases in which the provided technical advice is ineffective, determine the cause, and, thereby, improve the system.

Figure7 shows the simulation model. For the simulation, we need to define the input model and output log. We use the progress report of the students as the input model. TA-Bot calculates the severity depending on each of the progress reports and provides technical advice accordingly. We record the day on which the technical advice is provided, the technical advice provided, and the severity as the output model.

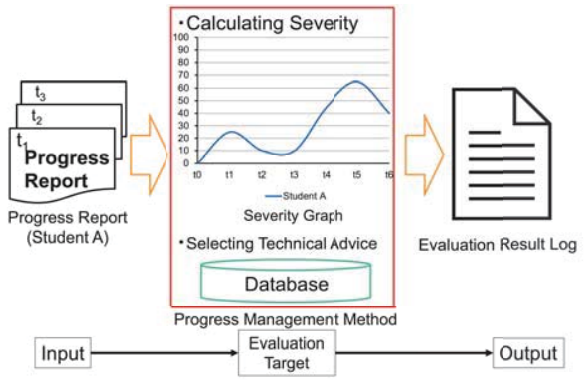


Figure 7: Simulation Model

Moreover, we need to define the developmental situation of each student when creating their progress reports. For this purpose, we define two student parameters, i.e., the level of programming skills of the students based on the difference in their programming proficiency and the development phase by dividing the development period into several stages.

We create the developmental situation by combining the two parameters and obtain the output log through simulation based on the developmental situation. Subsequently, we can confirm whether the technical advice is effective and how the severity varied by analyzing the output log and improve the system based on the result. In contrast to a real-world experiment, we can repeat the simulation to define new parameters and reconsider existing ones. Therefore, we can re-evaluate the system for the same situation after adjusting the timing and type of the technical advice.

4.1 Simulation Method

The evaluation method assumes that the developmental progress of every student is different. The objective of the class that is considered in this study is to implement a telephone switchboard using code, within a period of 6 months with the students submitting weekly progress reports. During the developmental period, TA-Bot uses the database to provide technical advice based on the progress report. Moreover, TA-Bot acquires experience in providing technical advice by learning from the advice in the database that was previously provided by the TA. Finally, it stores new information from the experiment to use as Static Advice. We analyze the result to output technical advice history.

4.2 Classifying students' programing skills

In software development, the status and timing of developmental delay and depend on the level of programming skills of the students. In other words, the cause of the delay is different for each student. Thus, different technical advice must be provided to different students. For this purpose, we classify programing skills into three levels.

1. Beginner
2. Intermediate
3. Advanced

The beginner level indicates that the student has finished basic programming training. However, they cannot write code without the aid of textbooks and web information.

The intermediate level indicates that the student can develop basic code without the aid of textbooks or web information. However, they need to collect information to perform applied programming.

The advanced level indicates that the student can provide technical advice to the Beginner and Intermediate level programmers. In addition, they can develop most software independently. Based on these three definitions, we can reproduce developmental delay situations, for which TA-Bot can provides technical advice depending on the skill level of the student.

4.3 Classifying development period

Furthermore, developmental progress is different for each student. In addition, the developmental situation and severity of the developmental delay depend on when the delay occurs. To reflect this, we divide the development period into three periods.

1. Phase1
2. Phase2
3. Phase3

Figure8 shows how the development period is divided. The phase "Early" denotes the period during which students create their plan and understand the program specifications. The "Middle" phase denotes the period during which the students write code. The "Late" phase denotes the period of the test phase

and extensions, if required. In addition, we divide degree of the delay into two types, viz., large and small. When the degree of the delay is large, the delay is considered to be extremely severe. However, when the degree of the delay is small, we don't judge that the delay is not severe. By dividing the development period in the aforementioned manner, we can reproduce eight different patterns and, consequently, several development situations. In addition, we can evaluate effectiveness of the advice provided in each period.

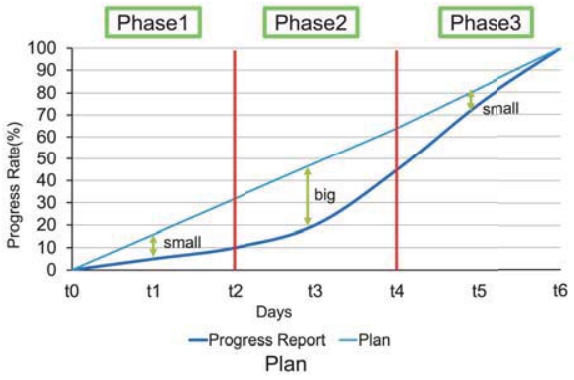


Figure 8: Classifying development period

4.4 Reproduction of development

A total of 24 cases of developmental situation can be reproduced by multiplying the three levels of programing skills with the eight types of development period. We create the progress reports based on these cases and evaluate the system using simulation. Therefore, we can reproduce typical cases.

4.5 Experiment

Figure9 shows a sample of the simulation. We define the parameters for each student, and included three students, studentA, studentB and studentC in each sample. We set the programing skill of studentA as Beginner, studentB as Intermediate, and studentC as Advanced. In the actual process of software development, each student has a different plan. Moreover, the technical advice provided by TA-Bot is not necessarily effective for all the students. Thus, we also set the effectiveness of the advice for each student. In the evaluation, the advice was considered as effective for A and B and ineffective for C. Furthermore, TA-Bot calculated the severity of delay for each student based on the progress report created based on the aforementioned parameters. Using the severity graph, we confirmed that the value change is different for each student. In other words, we were able to identify differences in the development situations by defining the parameters. Therefore, we can adjust the thresholds deciding the type of technical advice.

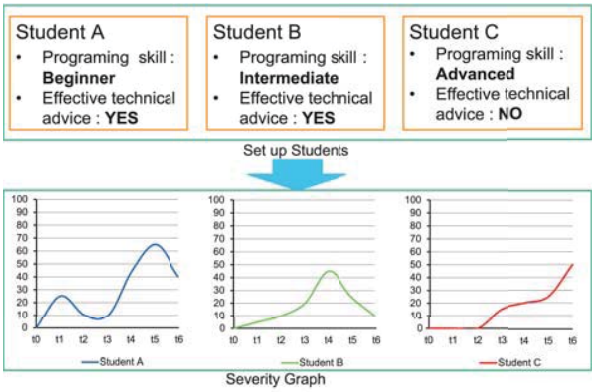


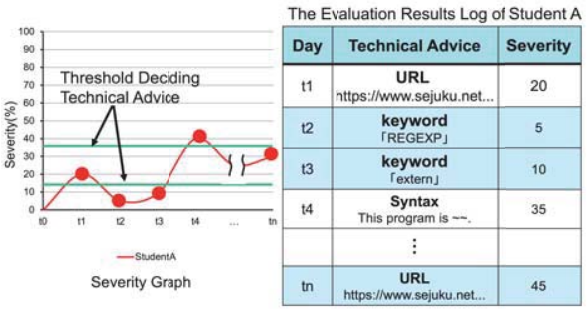
Figure 9: Experiment

4.6 Analyzing the evaluation result

We analyze the result of the evaluation using the severity graph, comment and output log.In other words, we analyze the timing providing technical advice is early or late by observing the severity, we adjust threshold deciding technicaln advice. The simulation outputs the technical advice, the day on which it was provided, and the severity of delay for each student. If the timing is early, we raise the threshold and make the timing slowly. If the timing is late, we lower the threshold and make the timing early. We orgnize two points how to judge the timing is early or slow.

1. The advice that student hasn't faced on task yet.
2. The advice that student has already solved.

We observe the severity after providing technical advice to judge these points. If the severity decrease after providing technical advice, we confirm the timing is appropriate. If the severity increase after providing technical advice, we confirm the timing is not appropriate. For confirming that, we analyze a comment in providing technical advice. The comment is written about students facing developmental tasks, so we are able to adjust the timing TA-Bot providing technical advice by investigating students' developmental situation based on the commnet. Therefore, we adjust the threshold deciding technical advie and technical advice by analyzing students facing developmental tasks. Figure10 shows a sample of output log. Based on the timing of the technical advice, we observed that the severity declined after the TA-bot provided the students with the URL of the website containing written development information. However, in t2, the severity rose after TA-Bot provided students with the keyword "REGEXP". In t3, the severity rose even further after TA-Bot provided students with the keyword "extern". In t4, the severity declined after TA-Bot provided the students with the programming syntax. Therefore, we can confirm that Student A found the syntax to be effective advice. Therefore, by analyzing the result logfile of the simulated experiment, we can confirm whether the technical advice provided for each student, depending on their development situation, was effective.



The Evaluation Results Log of Student A

Day	Technical Advice	Severity
t1	URL https://www.sejuku.net...	20
t2	keyword 「REGEXP」	5
t3	keyword 「extern」	10
t4	Syntax This program is ~~.	35
	⋮	
tn	URL https://www.sejuku.net...	45

Figure 10: Analyzing Logfile

5 Conclusion

In this study, we proposed a progress management method using TA-Bot for a software development class. To monitoring student progress, TA-Bot quantified the severity of software developmental delay and decided the timing of technical advice based on it. Consequently, we were able to change the thresholds deciding the type of technical advice.

For the experiment, we intentionally created students' development situations. We defined two parameters reflecting development situations, viz., programming skill levels and software development period. These two parameters were employed to faithfully reproduce development situations. The total number of cases that we could reproduce was 24. In the process of the simulated experiment, we were enabled to adjust the timing and effectiveness of the technical advice. This adjustment was enabled by the repeatability of the simulation.

In this study, the system was evaluated using simulations. However, the advice data is insufficient for properly evaluating the effectiveness of the technical advice. Thus, new technical advice must be added. Furthermore, we intentionally created development situations and repeated the evaluation. Thus, we could optimize the timing of the technical advice provided to individual students. However, actual development situations could be more complex. Consequently, the proposed method may not be able to effectively optimize the timing of the technical advice. Therefore, in future work, we will evaluate the system not only using simulation, but also using real development situations with the help of student participation. For this purpose, we will extend the system and modify it.

REFERENCES

[1] Miyuki Yamamoto, Koichi Fukuoka, Ryozo Kiyohara, Yoshiaki Terashima. "Progress Management Method for Software Development Project-based Learning Using Automated Teaching Assistants" 2019 Twelfth International Conference on Mobile Computing and Ubiquitous Network(ICMU). IEEE, pp. 52–55 (2019)

[2] https://www.enpit.jp/index.html

[3] Slack ,https://slack.com/intl/ja-jp/

[4] T. Zimmermann, J. Coolen, J. Gross, P. -M. Pedrot and G. Gilbert, "Advantages of maintaining a multi-task project-specific bot: an experience report," in IEEE Software, doi: 10.1109/MS.2022.3179773.

[5] I. Beschastnikh, M. F. Lungu and Y. Zhuang, "Accelerating Software Engineering Research Adoption with Analysis Bots," 2017 IEEE/ACM 39th International Conference on Software Engineering: New Ideas and Emerging Technologies Results Track (ICSE-NIER), 2017, pp. 35-38, doi: 10.1109/ICSE-NIER.2017.17.

[6] S. Saiki et al., "A Study of Practical Education Program on AI, Big Data, and Cloud Computing through Development of Automatic Ordering System," 2018 IEEE International Conference on Big Data, Cloud Computing, Data Science & Engineering (BCD), 2018, pp. 31-36, doi: 10.1109/BCD2018.2018.00013.

An Experiment of an Office Worker Passing by the Surrogate Robot of a Remote Worker Using VR Video

Kosuke Sasaki†, Zijie Yuan†and Tomoo Inoue‡

†Graduate School of Library, Information and Media Studies, University of Tsukuba, Japan
‡Faculty of Library, Information and Media Science, University of Tsukuba, Japan
{ksasaki, inoue}@slis.tsukuba.ac.jp

Abstract - This study focuses on a situation in which a local worker and a surrogate telepresence robot of a remote worker pass by in an office corridor or passageway. A method of giving the remote worker's image onto the telepresence robot using VR technology has been proposed. In this paper, we examined if the method was useful for workers through their subjective evaluation. The result of a questionnaire survey and an interview indicated that the workers using this system could predict where the remote worker was moving to by looking at his/her upper body and could avoid collision.

Keywords: Cooperative work, Positional relationship, Nonverbal cue, Telepresence robot

1 Introduction

Telework has increased in recent years, and there are more and more situations in which robots and people work cooperatively. In particular, this study focuses on work environment using mobile telepresence robots in which a remote worker operates the robot and works with a local worker. When a telepresence robot and a human are in the same space, it is important for them to be aware of each other's position. Considering a situation where a worker and a robot pass each other in an office corridor or passageway, it is necessary to know how the partner will move not to bump into each other.

Inoue and Yuan have proposed a method to give a human image to a telepresence robot using VR technology to address this problem[1]. However, the experiment environment was not so realistic in that the initial distance between a human and a robot was too close. Also, the paper only reported an initial evaluation and did not focus on the interaction between the worker and the robot. In this paper, we evaluate this method based on subjective evaluations obtained from participants.

In the experiment conducted in this study, workspaces at two remote locations were prepared. In each workspace, a mobile robot equipped with an RGB camera was placed. Each robot moves synchronizing the position of its respective remote worker. Each camera attached to the robot gave the partner's image to an HMD which respective workers wore. In this environment, workers were paired up and performed a passing-by task that simulated passing each other in a company corridor. The results suggest that if the workers can see their partner through the HMD, they move to avoid collision predicting the direction of the partner by looking at their upper body.

2 Related Works

Passing by a human and a robot in office environment, which is the situation this study focuses on, can be regarded as remote cooperative work between a local worker and a remote worker. We first describe remote cooperative work in environment using fixed displays.

2.1 Remote Cooperative Work Using Fixed Displays

Unlike face-to-face environments, visual nonverbal information such as gaze information or body movement may lack in remote environment. Since nonverbal information is important for smooth communication, studies have been conducted to cooperate between remote locations while communicating visual nonverbal information to each other.

Making workers feel as if they are in the same space even though they are in remote environment is one method of supporting remote cooperative work. For example, some systems use a common workspace or background for feeling that being in the same space[2], [3] or a system that overlays the remote partner on a landscape behind the display[4]. However, the use of such fixed displays places significant restrictions on the worker's physical movement, making them unsuitable for cooperative work that involves physical movement. Therefore, some systems use AR or VR technologies to remove the restrictions.

2.2 Remote Cooperative Work Using AR or VR Technologies

Systems using AR or VR technologies with HMDs have also been proposed to solve the problem of limited body movement. For example, Fuchs et al. and You et al. have proposed systems that support remote cooperative work by immersing the user in the same virtual space with a remote partner[5], [6].

Using avatars also seems to work for smooth remote collaborative work. In fact, a number of studies have reported the effectiveness[7]–[10]. AR and VR technologies can reduce the restrictions on worker movement, however, the worker does not have a physical body and cannot interact with the real space in virtual spaces where many studies use in their system. Furthermore, using avatars can miss non-verbal information of workers before they are made into avatars.

2.3 Remote Cooperative Work Using Telepresence Robots

Telepresence robots have the possibility to solve the above issues, that is, reduce the limitation of body movement and enable workers to interact with real space or objects without missing nonverbal information. MeBot[11] developed by Adalgeirsson et al. and iRIS[12] developed by Kawanobe et al. can transmit gaze information and head movement of a partner by moving the robot's head and the display attached to the head. Such technology may be effective for cooperative work in remote work. For instance, in a remote meeting, these technologies made it easier to identify the conversation partner and reduced time required to answer questions.

Another study has suggested a system that immerses the remote worker in the robot which can display the remote worker's face to local workers and perform the same arm movement of the remote worker[13], [14]. Such systems lead remote workers to collaborate with local workers.

A mobile robot can present remote workers' position in space to local workers, enabling real-space interaction and cooperative work between remote locations without restrictions on movement within the space. Third Eye[15] and Yang et al. studies[16] used a one-directional but immersive mobile robot for remote cooperative work. However, the remote worker's movement are limited to those that the robots can move, which may miss some nonverbal information.

Based on these studies, we decided to use VR technology and give the mobile telepresence robot the appearance of the partner as is. This makes it possible to convey visual information to the local worker.

3 Environment of the System

This study used a system based on the previous study[1]. Figure 1(a) shows the environment of the system, which a worker and a telepresence robot pass by each other. Two workspaces were prepared and a pair of a worker and a telepresence robot was arranged in each workspace. Telepresence robots equipped with a 360-degree camera (as shown in Figure 1(b)) were placed in front of each worker in different workspaces. The robots moved synchronizing the respective worker's positions in each workspace.

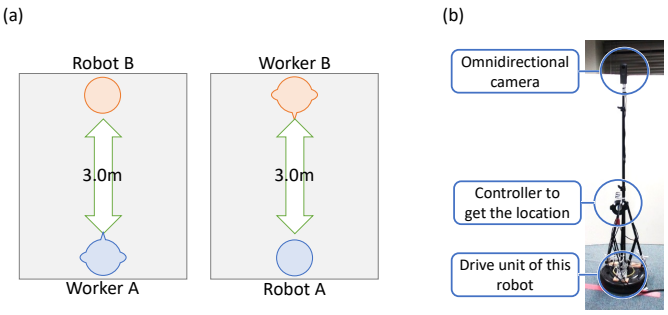


Figure 1: (a) Environment of the experiment / (b) Appearance of the telepresence robot

Both workers wear HMDs, and they could see images from the robot's camera placed in front of the partner. In this way, an environment was created in which the local worker and the telepresence robot that surrogates the remote worker pass by each other.

In this study, VIVE™ by HTC Corporation[1] was used for the HMD worn by the worker and for the controller and tracking camera to acquire the coordinates of the robot and worker. The iRobot® Create 2 by iRobot[2] was also used for the driving part of the robot. A height-adjustable tripod for a camera was placed on the robot, and a VIVE™ controller was attached to acquire the robot's coordinates. Additionally, a RICHO's[3] 360-degree camera, Theta V™, was mounted at the top of the tripod to capture the partner's appearance.

4 Experiment

4.1 Passing-by Task

In this study, we conducted an experiment of passing-by task in which a telepresence robot and a remote worker passed by each other (approved by the Ethics Review Committee of Faculty of Library, Information and Media Science, University of Tsukuba (Notification number: 20-15)). In this experiment, the worker and robot initially stood facing each other at a distance of 3.0 meters as shown in Figure 1(a). On the experimenter's cue, they started walking, the worker aiming at the robot's initial position and the robot aiming at the worker's initial position.

4.2 Experiment Conditions

We compared the following two conditions; the bidirectional condition and the unidirectional condition. The bidirectional condition was that workers were immersed in robots in each

[1]VIVE - VR Headsets, Games, and Metaverse Life, https://www.vive.com/ (Visited on Dec 10, 2022)

[2]Coding Robots, Learning Library & STEM Outreach | iRobot Education, https://edu.irobot.com/ (Visited on Dec 10, 2022)

[3]Ricoh Global | EMPOWERING DIGITAL WORKPLACES, https://www.ricoh.com/ (Visited on Dec 10, 2022)

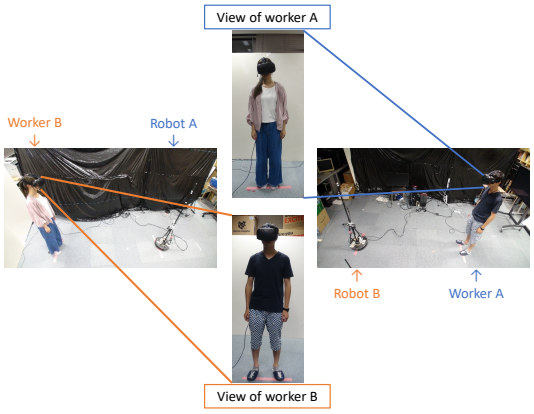


Figure 2: The bidirectional condition

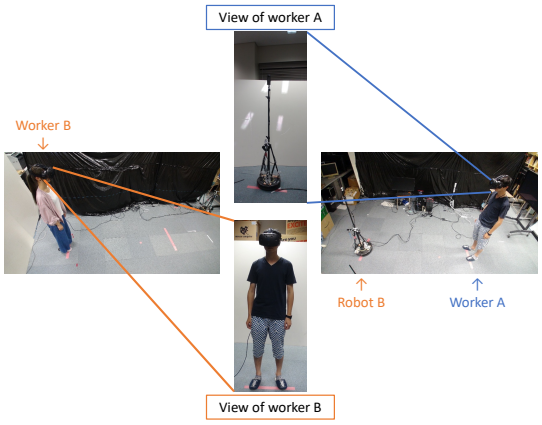


Figure 3: The unidirectional condition

other's remote workspace. Workers wore HMDs at both remote locations, and each HMD displays the video from the camera mounted on the robot at the remote location as shown in Figure 2. On the other hand, the unidirectional condition was a condition in which only one worker was immersed in a robot in the remote workspace. Only one worker was immersed in the remote robot and only took the role of the controller of the robot. Another worker also wore an HMD, but it displayed the see-through image in front of him/her as it is. Also, a robot in front of the worker who was a controller of the remote robot was not placed as shown in Figure 3).

In both conditions, workers were not allowed to talk to each other in order to examine the visual information transfer between workers' actions. Therefore, during the experiment, workers moved based only on visual information.

4.3 Participants

Six pairs consisting of 12 graduate students (all male) participated in the experiment. Of the 12 participants, five used an HMD for the first time in this experiment. Only one of the six pairs had never met each other before.

4.4 Procedure

The experiment was conducted in a within-participant design. Each pair participate both of the conditions with counterbalanced. In each condition, participants practiced the passing-by task twice to familiarize themselves with the task using the telepresence robot. Also, the experimenter explained the task until both participants understood.

After that, the real task is performed once. After the real task, the sequence was repeated under different conditions. The unidirectional condition was asymmetry as opposed to the bidirectional condition. Therefore, after completing the real task once in the unidirectional condition, workers changed roles and performed the practice and real tasks again. Therefore, the real task has performed a total of six times (one time for each pair) for all participants in the bidirectional condition and 12 times (two times for each pair) in the unidirectional condition.

For participants' safety, one experimenter was assigned to each workplace to prevent workers from colliding with the robot and catching their feet on the cords connecting the robot and the processing PC during the task.

After the real task in each condition, a questionnaire survey was conducted. We set up 10 questions for the questionnaire, asking not only about the robot's usability but also about the presence of the partner and the sense of being in the same room with the robot. Participants responded to the evaluation items on a 7-point scale from "strongly disagree" to "strongly agree." We also conducted an unstructured interview after the completion of the task in both conditions, and participants were asked to respond freely to questions about their impressions of the task and the system.

5 Result and Discussion

Ratings were treated as a score from 1 to 7, and a Wilcoxon signed-rank test was conducted for each question item. The results are shown in Table1. In particular, the scores following four questions for the bidirectional condition were significantly higher than those for the unidirectional condition.: Q4 "I could clearly know the motion of the partner." ($p < .01$), Q7 "I found the movement and direction of the partner." ($p < .01$), Q8 "I could predict where the partner was moving to." ($p < .01$), and Q9 "Partner's movement looked realistic." ($p < .05$).

In the interview, we obtained comments such as "I could predict the robot's path when looking at a human (P3: Bidirectional condition)," "I could clearly see its direction of movement when looking at a human (P5: Bidirectional condition)," "I could not see the robot's direction of movement (P2: Unidirectional condition)," "I could not see the robot's direction of movement (P7: Unidirectional condition)," and "I was confused about the robot's direction of movement (P8: Unidirectional condition)."

These results of the questionnaire and the interview indicate that workers predicted the robot's direction of movement, i.e.

Table 1: The results of the questionnaire. "Bi" and "Uni" means the median of rating in the "Bidirectional condition" or "Unidirectional Condition" respectively. p-values are the results of Wilcoxon signed-rank test.

(** $p < .01$, * $p < .05$, † $p < .10$)

	Items	Bi	Uni	p-value
Q1	After using the system, I felt uncomfortable similar to motion sickness	5	5	.79
Q2	The video was clear and pretty.	5	4.5	.08†
Q3	I could feel like the remote person was in the same room with me.	6.5	6	.13
Q4	I could clearly know the motion of the partner.	6	5	.008**
Q5	I felt like I passed the other person in the same room.	5.5	5.5	.69
Q6	The field of view changed naturally according to my movement.	5.5	5.5	1.00
Q7	I found the movement and direction of the partner.	6	4.5	.003**
Q8	I could predict where the partner was moving to.	6	4	.001**
Q9	Partner's movement looked realistic.	6	4.5	.02*
Q10	The experience in the virtual environment was consistent with experiences lived in real life.	5.5	5	.16

the partner's direction of movement, when the partner's image was displayed on their HMDs.

The importance of partners' image in collaborative work has been proven in previous studies, too.

In a study of a collaborative bicycle repair task in face-to-face and in remote found the visual cues in the partner's image could make the grounding in their conversation and the task[17].

Another study of a remote collaborative task of assembling a toy was conducted in a specially designed space where the local worker and a table were surrounded by eight fixed displays showing a remote worker. It found the local worker could predict the remote worker's movement by seeing the upper body of the partner and could prepare for the next action[18].

In contrast to these studies on predefined collaborative work in a fixed place, our study indicates the usefulness of showing the partner's video to pass by the robot smoothly by predicting its moving direction.

6 Conclusion

In this study, we focused on a situation in which a telepresence robot and a local worker pass by in an office corridor or passageway. An VR system that displays video of a remote worker at the position of a surrogate telepresence robot to the local worker to avoid collision has been studied[1]. However, its evaluation was very preliminary and was not very realistic. In this paper, we experimented with a passing-by task and conducted a questionnaire survey and an interview that asked about workers' feelings during the task. The results suggested that presenting the partner's upper body was used to predict the partner's movement, and thanks to this, the worker can avoid collision with the surrogate telepresence robot.

ACKNOWLEDGEMENT

The authors would like to thank Dr. Yasuhito Noguchi for his cooperation in conducting the experiment.

REFERENCES

[1] T. Inoue and Z. Yuan, "Remote video figure achieves smooth cooperative movement in a bidirectional telepresence robot environment," in *Collaboration and Technology*, A. Rodrigues, B. Fonseca, and N. Preguiça, Eds. Cham: Springer International Publishing, 2018, pp. 91–104.

[2] J. C. Tang and S. Minneman, "Videowhiteboard: Video shadows to support remote collaboration," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '91. New York, NY, USA: Association for Computing Machinery, 1991, pp. 315–322. [Online]. Available: https://doi.org/10.1145/108844.108932

[3] L. Handberg, C. Gullströ, J. Kort, and J. Nyström, "Spatial and social connectedness in web-based work collaboration," in *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*, ser. CSCW '16 Companion. New York, NY, USA: Association for Computing Machinery, 2016, pp. 45–48. [Online]. Available: https://doi.org/10.1145/2818052.2874321

[4] T. Inoue, M. Nawahdah, and Y. Noguchi, "User's communication behavior in a pseudo same-room videoconferencing system bhs," *International Journal of Informatics Society (IJIS)*, vol. 6, pp. 39–47, 11 2014.

[5] H. Fuchs, A. State, and J. Bazin, "Immersive 3d telepresence," *Computer*, vol. 47, no. 7, pp. 46–52, July 2014.

[6] B.-J. You, J. R. Kwon, S.-H. Nam, J.-J. Lee, K.-K. Lee, and K. Yeom, "Coexistent space: Toward seamless integration of real, virtual, and remote worlds for 4d+ interpersonal interaction and collaboration," in

SIGGRAPH Asia 2014 Autonomous Virtual Humans and Social Robot for Telepresence, ser. SA '14. New York, NY, USA: Association for Computing Machinery, 2014. [Online]. Available: https://doi.org/10.1145/2668956.2668957

[7] G. A. Lee, T. Teo, S. Kim, and M. Billinghurst, "Mixed reality collaboration through sharing a live panorama," in *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*, ser. SA '17. New York, NY, USA: Association for Computing Machinery, 2017. [Online]. Available: https://doi.org/10.1145/3132787.3139203

[8] T. Piumsomboon, A. Day, B. Ens, Y. Lee, G. Lee, and M. Billinghurst, "Exploring enhancements for remote mixed reality collaboration," in *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*, ser. SA '17. New York, NY, USA: Association for Computing Machinery, 2017. [Online]. Available: https://doi.org/10.1145/3132787.3139200

[9] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst, "Mini-me: An adaptive avatar for mixed reality remote collaboration," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY, USA: Association for Computing Machinery, 2018. [Online]. Available: https://doi.org/10.1145/3173574.3173620

[10] T. Teo, A. F. Hayati, G. A. Lee, M. Billinghurst, and M. Adcock, "A technique for mixed reality remote collaboration using 360 panoramas in 3d reconstructed scenes," in *25th ACM Symposium on Virtual Reality Software and Technology*, ser. VRST '19. New York, NY, USA: Association for Computing Machinery, 2019. [Online]. Available: https://doi.org/10.1145/3359996.3364238

[11] S. O. Adalgeirsson and C. Breazeal, "Mebot: A robotic platform for socially embodied telepresence," in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, March 2010, pp. 15–22.

[12] H. Kawanobe, Y. Aosaki, H. Kuzuoka, and Y. Suzuki, "iris: A remote surrogate for mutual reference," in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, March 2013, pp. 403–403.

[13] S. Tachi, "Telexistence: Enabling humans to be virtually ubiquitous," *IEEE Computer Graphics and Applications*, vol. 36, no. 1, pp. 8–14, Jan 2016.

[14] S. Tachi, K. Watanabe, K. Takeshita, K. Minamizawa, T. Yoshida, and K. Sato, "Mutual telexistence surrogate system: Telesar4 - telexistence in real environments using autostereoscopic immersive display -," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sep. 2011, pp. 157–162.

[15] A. P. Tarun, N. M. Baig, J. S.-K. Chang, R. Tanvir, S. Shihipar, and A. Mazalek, "Third eye: Exploring the affordances of third-person view in telepresence robots," in *Social Robotics*, M. A. Salichs, S. S. Ge, E. I. Barakova, J.-J. Cabibihan, A. R. Wagner, Á. Castro-González, and H. He, Eds. Cham: Springer International Publishing, 2019, pp. 707–716.

[16] L. Yang, B. Jones, C. Neustaedter, and S. Singhal, "Shopping over distance through a telepresence robot," *Proc. ACM Hum.-Comput. Interact.*, vol. 2, no. CSCW, Nov. 2018. [Online]. Available: https://doi.org/10.1145/3274460

[17] S. R. Fussell, R. E. Kraut, and J. Siegel, "Coordination of communication: Effects of shared visual context on collaborative work," in *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, ser. CSCW '00. New York, NY, USA: Association for Computing Machinery, 2000, pp. 21–30. [Online]. Available: https://doi.org/10.1145/358916.358947

[18] N. Yamashita, H. Kuzuoka, K. Hirata, S. Aoyagi, Y. Shirai, K. Kaji, and Y. Harada, "Effects of showing user's upper body in video-mediated collaboration," *IPSJ Journal*, vol. 51, no. 4, pp. 1152–1162, apr 2010. [Online]. Available: https://ci.nii.ac.jp/naid/110007970719/en/

Keynote Speech
(Chair: Tomoyuki Yashiro)

TOSHIBA

Toward Software Defined Social Infrastructure based on Cyber Physical Systems Technology

Information & Communication
Platform Laboratories,
Toshiba R&D Center

Takeshi Saito

September 2, 2022

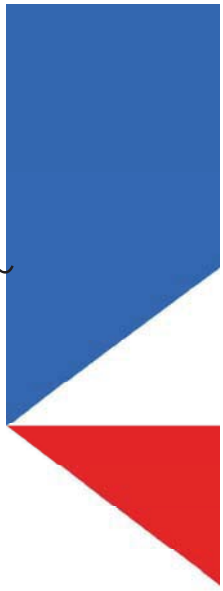
Self Introduction

Takeshi Saito (斉藤　健)

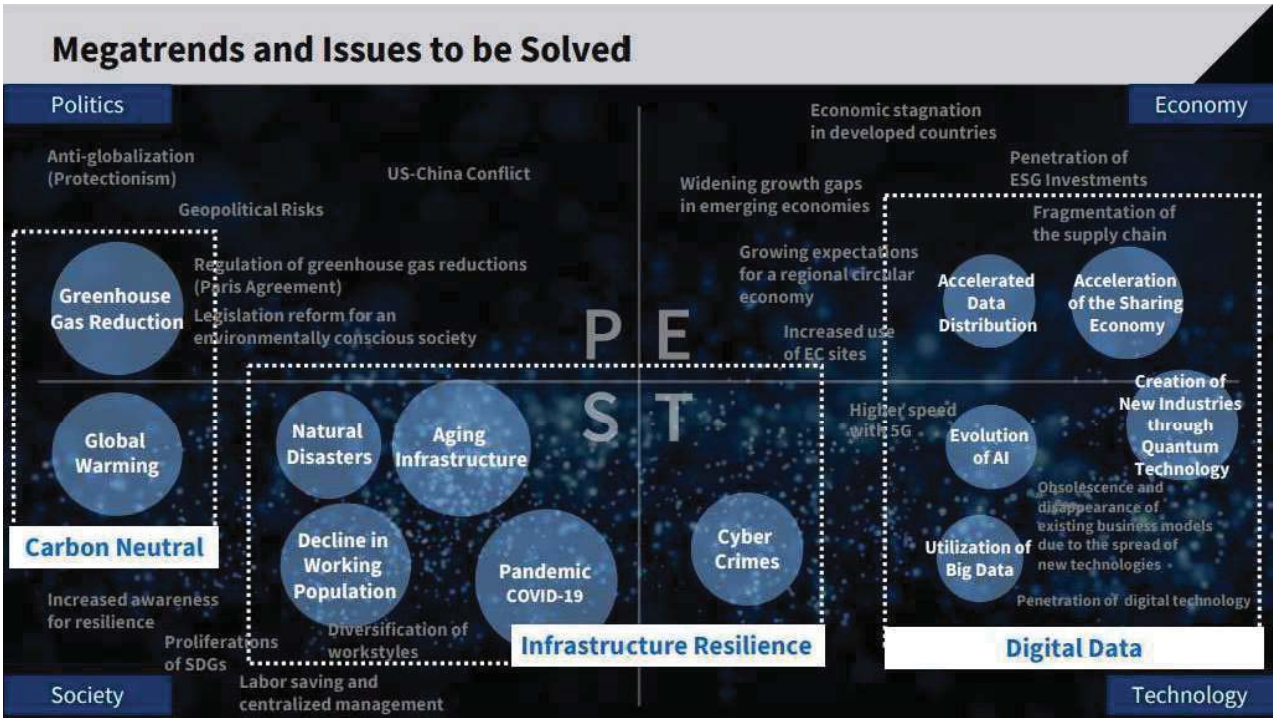
- Joined Toshiba Corporation in 1989
- Experienced in research institutes, home appliance divisions, and social infrastructure divisions
- Engaged in the application of Internet technology
- Currently in charge of information and communication platforms at the R&D Center
  + Circuits, wireless, network, computing, IoT edge, cloud, video system, distributed power supply system
- Representative Director of "ECHONET Consortium" from 2022

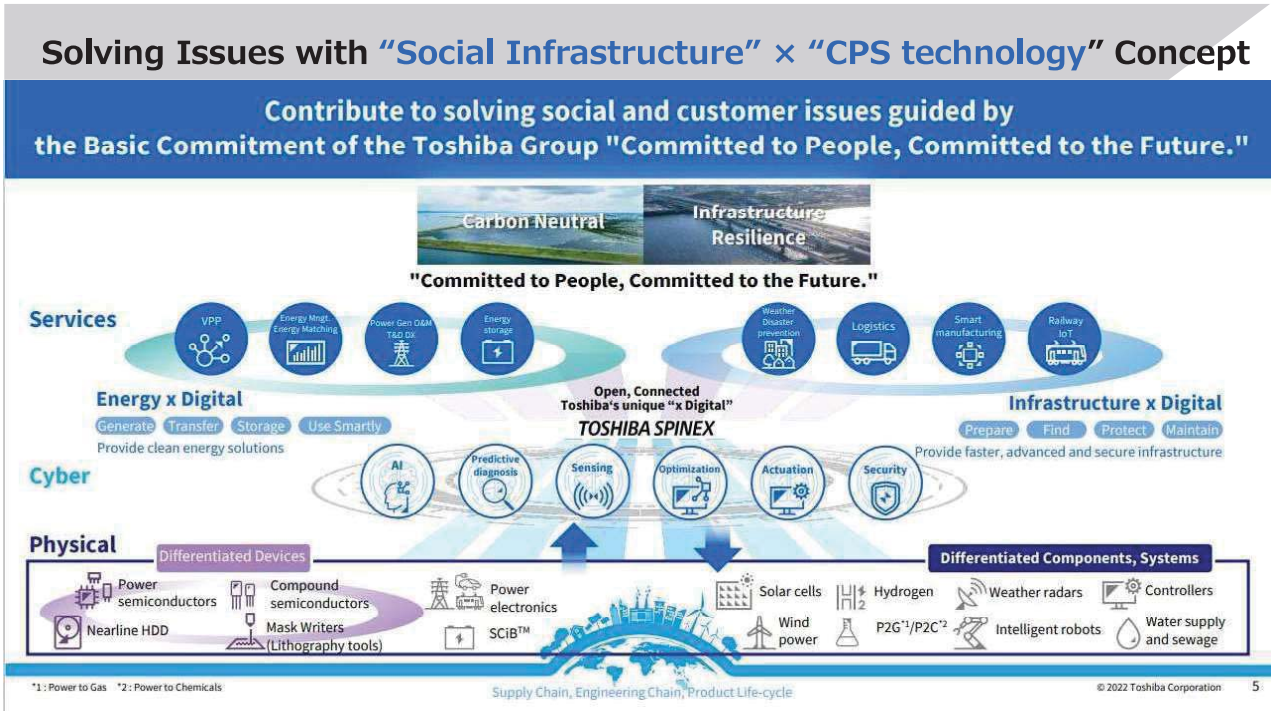
 2

Contents

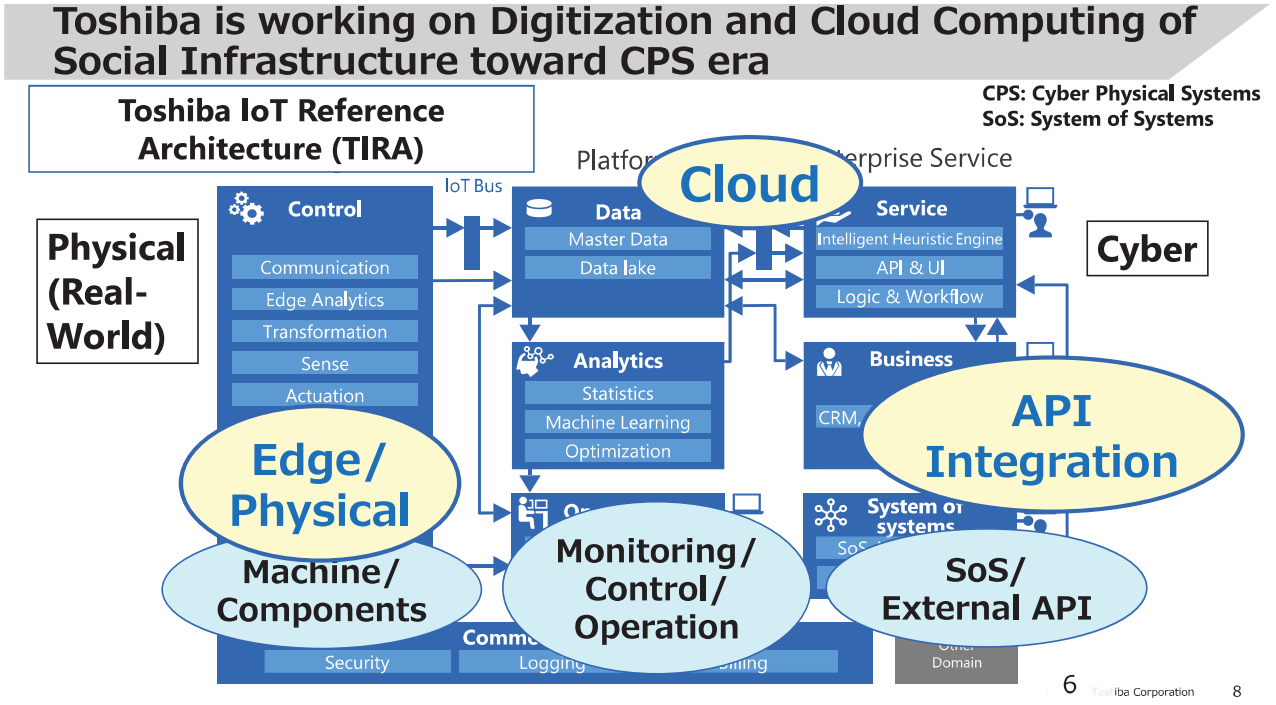


01 Converting Social Infrastructure to CPS(Cyber Physical Systems)

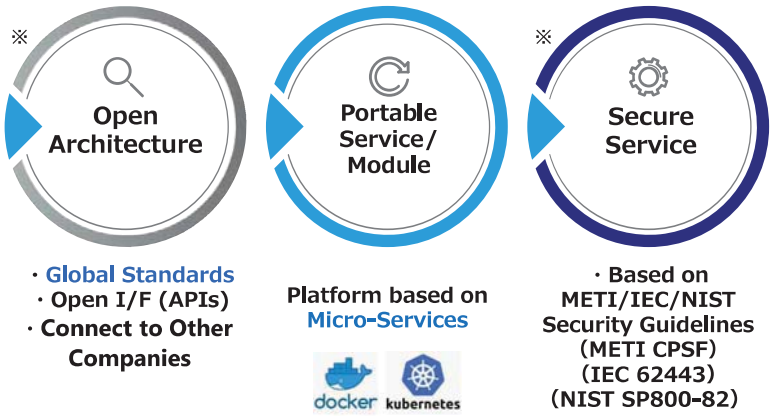

Megatrends and Issues to be Solved
Politics
Economy
Anti-globalization (Protectionism)
US-China Conflict
Geopolitical Risks
Economic stagnation in developed countries
Penetration of ESG Investments
Widening growth gaps in emerging economies
Fragmentation of the supply chain
Growing expectations for a regional circular economy
Regulation of greenhouse gas reductions (Paris Agreement)
Legislation reform for an environmentally conscious society
Greenhouse Gas Reduction
Accelerated Data Distribution
Acceleration of the Sharing Economy
P E
Increased use of EC sites
S T
Creation of New Industries through Quantum Technology
Higher speed with 5G
Evolution of AI
Global Warming
Natural Disasters
Aging Infrastructure
Obsolescence and disappearance of existing business models due to the spread of new technologies
Utilization of Big Data
Carbon Neutral
Decline in Working Population
Pandemic COVID-19
Cyber Crimes
Penetration of digital technology
Increased awareness for resilience
Proliferations of SDGs
Diversification of workstyles
Infrastructure Resilience
Digital Data
Society
Labor saving and centralized management
Technology


Solving Issues with "Social Infrastructure" × "CPS technology" Concept
Contribute to solving social and customer issues guided by the Basic Commitment of the Toshiba Group "Committed to People, Committed to the Future."
Carbon Neutral
Infrastructure Resilience
"Committed to People, Committed to the Future."
Services
VPP
Energy Mngt. Energy Matching
Power Gen O&M T&D DX
Energy storage
Weather Disaster prevention
Logistics
Smart manufacturing
Railway IoT
Energy x Digital
Generate
Transfer
Storage
Use Smartly
Provide clean energy solutions
Open, Connected
Toshiba's unique "x Digital"
TOSHIBA SPINEX
Infrastructure x Digital
Prepare
Find
Protect
Maintain
Provide faster, advanced and secure infrastructure
Cyber
AI
Predictive diagnosis
Sensing
Optimization
Actuation
Security
Physical
Differentiated Devices
Power semiconductors
Compound semiconductors
Nearline HDD
Mask Writers (Lithography tools)
Power electronics
SCiB™
Differentiated Components, Systems
Solar cells
Hydrogen
Weather radars
Controllers
Wind power
P2G*1/P2C*2
Intelligent robots
Water supply and sewage
*1 : Power to Gas *2 : Power to Chemicals
Supply Chain, Engineering Chain, Product Life-cycle
© 2022 Toshiba Corporation
5


Any Social Infrastructures are our CPS targets
CPS: Cyber Physical Systems
Social Infrastructure Solutions
Components
Solar Power
Wind Power Generation
Smart Meter
Water Infrastructure
Railway Infrastructure
Communication/ Broadcasting Infrastructure
Highway Infra.
Substation
Hydrogen System
Distribution
Air Conditioning
Elevator
Lighting
Factory Solution
Building Solution
Weather Infrastructure
Industrial
Automotive
Data Centers
Power Electronics
Sensing
HDD
Battery
Motors
POS System
We will respond to the huge social trends and to the various needs of customers/regions with CPS/Digitalization.


Toshiba is working on Digitization and Cloud Computing of Social Infrastructure toward CPS era
Toshiba IoT Reference Architecture (TIRA)
CPS: Cyber Physical Systems
SoS: System of Systems
Cloud
IoT Bus
Physical (Real-World)
Cyber
Control
Communication
Edge Analytics
Transformation
Sense
Actuation
Data
Master Data
Data lake
Service
Intelligent Heuristic Engine
API & UI
Logic & Workflow
Analytics
Statistics
Machine Learning
Optimization
Business
CRM,
Edge/ Physical
API Integration
Machine/ Components
Monitoring/ Control/ Operation
SoS/ External API
Security
Domain
6
8

TIRA: Basic Design Policy

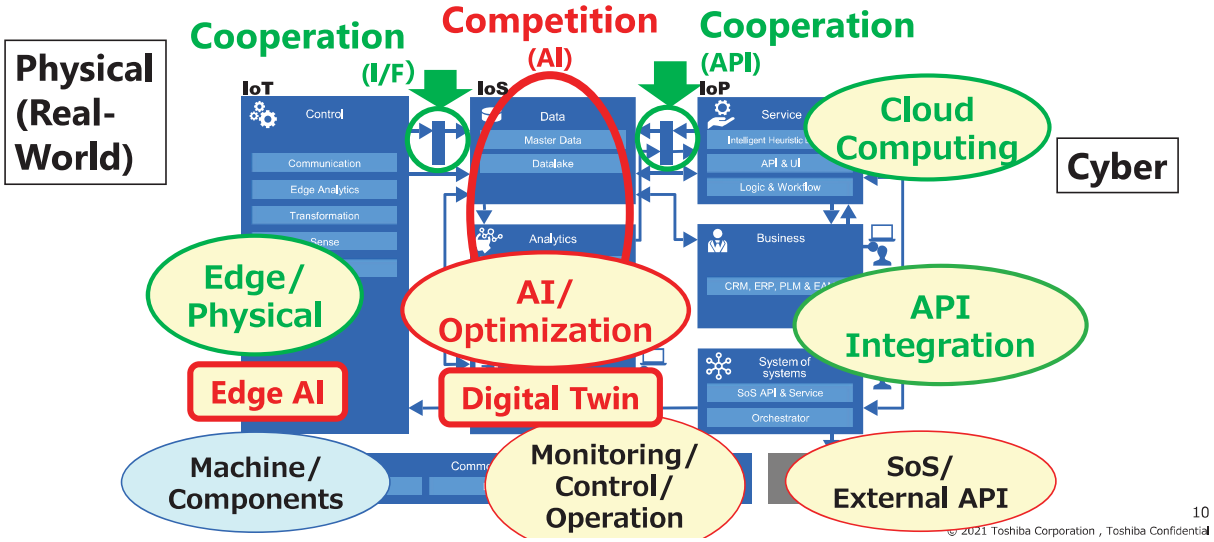


 9

Cooperation & Competition on CPS-based Social Infrastructure

CPS: Cyber Physical Systems

Cooperation with OpenAPI & Open I/F.
Differentiation (Competition) through AI/Optimization



 10

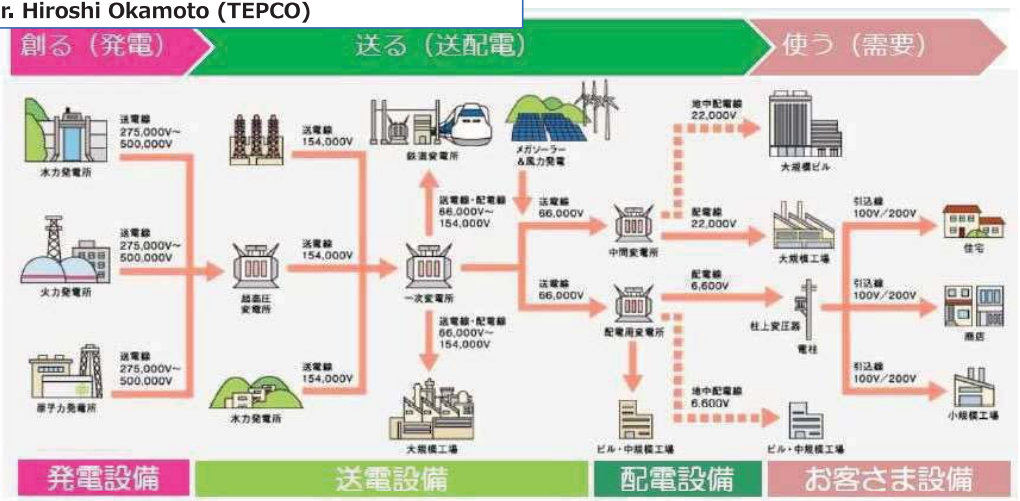
02 Software Defined Social Infrastructure

～Examples in Distributed Energy System Area～

 11

What is happening in the Energy Infrastructure toward Carbon Neutral?

Ministry of the Environment meeting materials
By Dr. Hiroshi Okamoto (TEPCO)

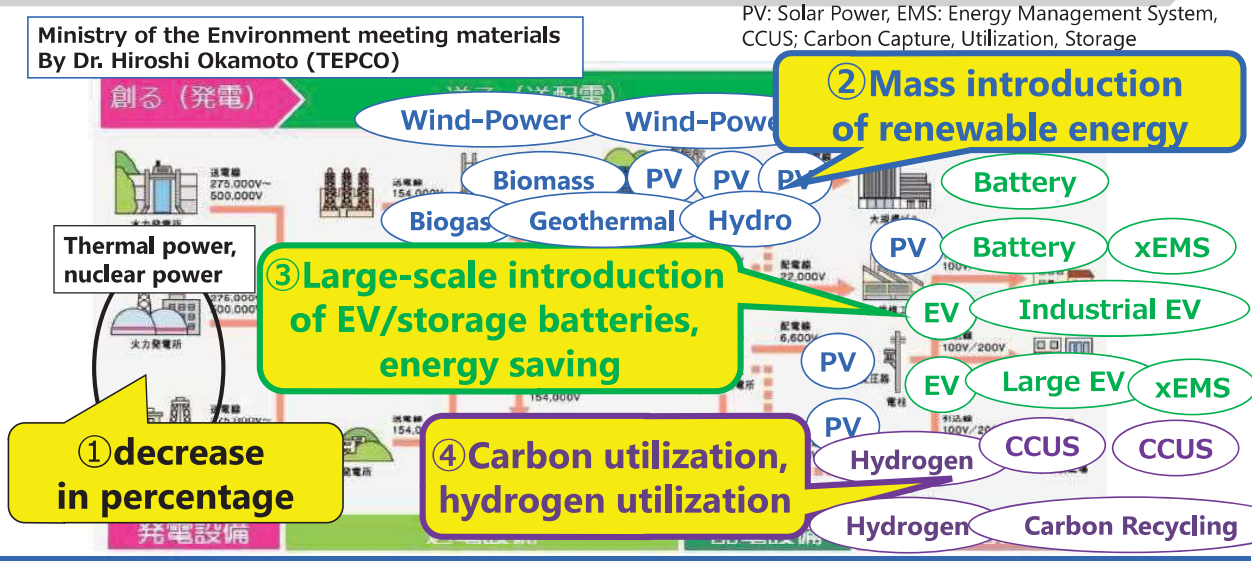


Conventional Energy Infrastructure

2

Key Technologies/New Technologies for Carbon Neutral

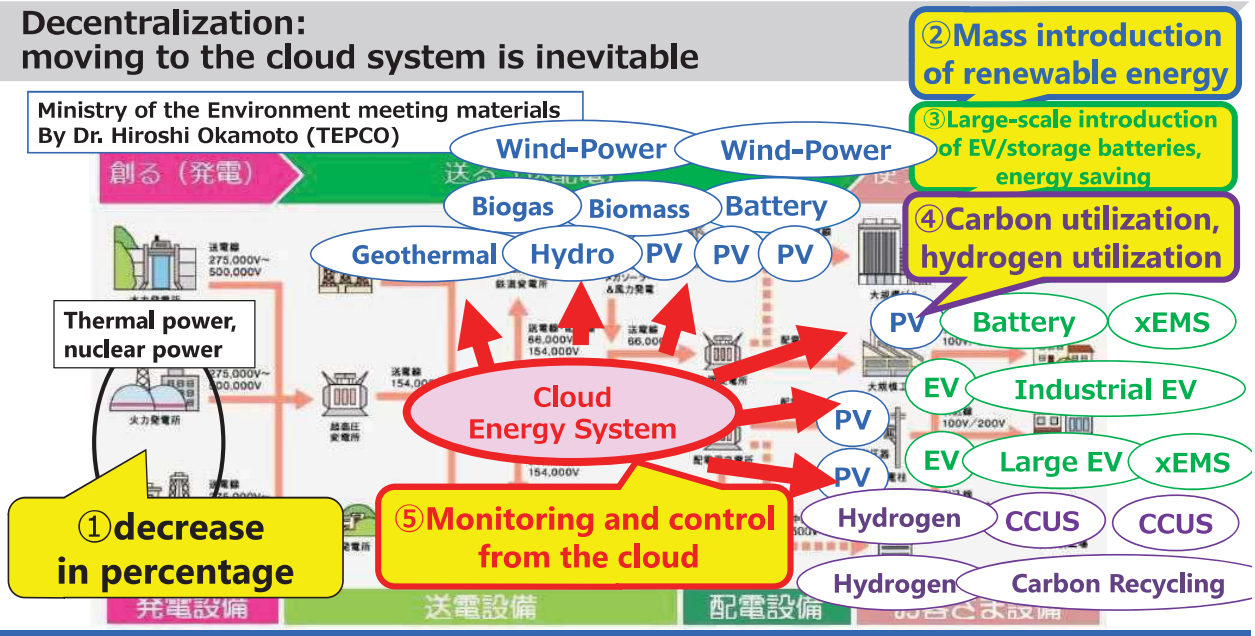
PV: Solar Power, EMS: Energy Management System,
CCUS; Carbon Capture, Utilization, Storage



Renewable energy, storage, and new technologies are deployed midstream to downstream as distributed power sources

3

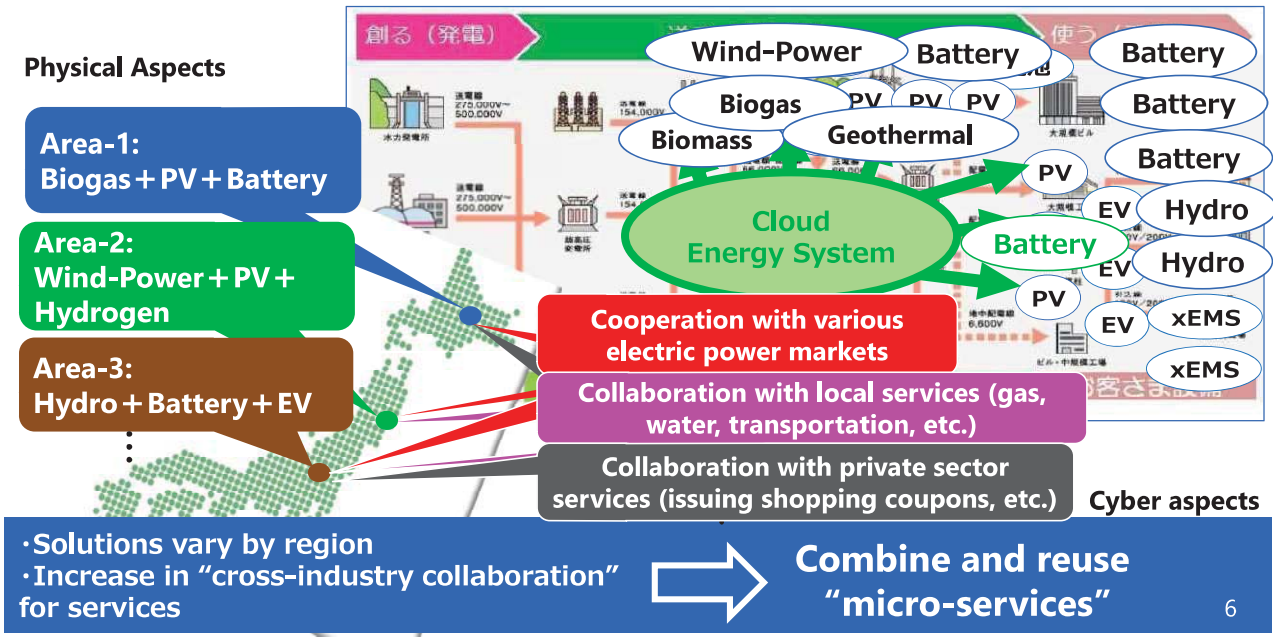
Decentralization: moving to the cloud system is inevitable



Cloud computing is the only solution for monitoring and controlling distributed energy resources that are diverse and spread out over a wide area.

5

Solutions and Services Provided Vary by Region and Operator



・Solutions vary by region
・Increase in “cross-industry collaboration” for services

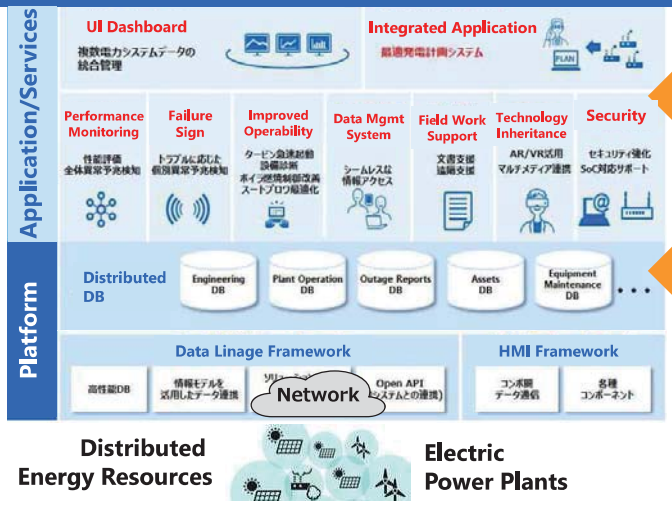
⇨ Combine and reuse “micro-services”

6

Energy System Solution (Preparing for Micro-Services on the Cloud)

We provide a combination of services to meet the various needs of each customer and region. A platform for providing “evolving operations and services” for power plants and VPPs

Implement Carbon Neutral Solutions on the Cloud



Micro-Services
（Can Choose Necessary Services）

API Integration
（Integration with Regions/ Other Companies）

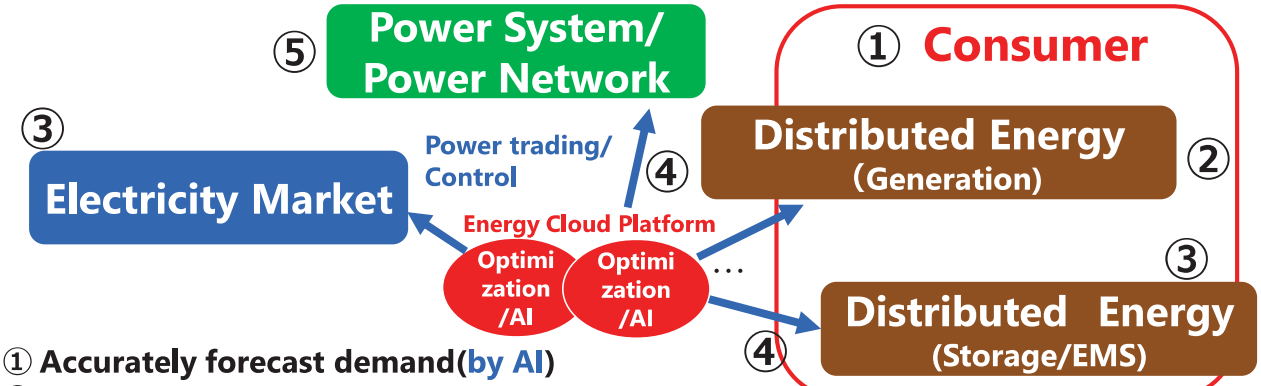
Distributed Energy Resources

Electric Power Plants

 16

Advanced Software Coordination between Markets/Grids/Distributed Energy Systems

Optimization/AI technology (software) achieve both the most Economical Operation (Profit Maximization, Digital) and the Grid Stabilization (Physical)



① Accurately forecast demand(by AI)
② Accurate prediction of renewable energy power generation(by AI)
③ Optimize the most economical way to procure power shortages(by AI)
④ Power trading/Control power system/physical power transmission(by IoT)
⑤While maintaining power system stability(by Power System Operation Tech.)

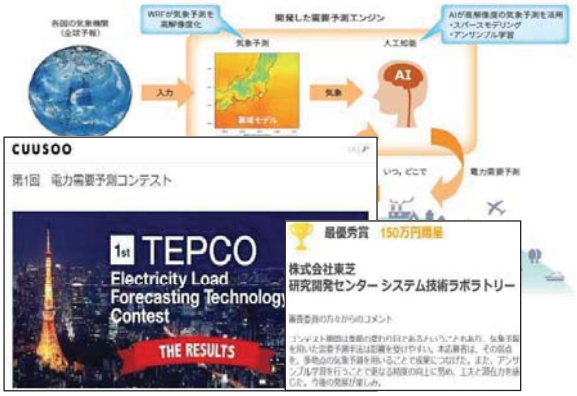
EMS:Energy Management System
DR: Demand Response

 17

Advanced Optimization/AI: Power Demand Forecasting / Photovoltaic Power Generation Forecasting

Combining our proprietary weather prediction system and machine learning methods to achieve highly accurate forecasting

Demand Forecasting



※東京電力HD様HP：http://www.tepco.co.jp/press/release/2017/1463817_8706.html ,

PV Power Generation Forecasting

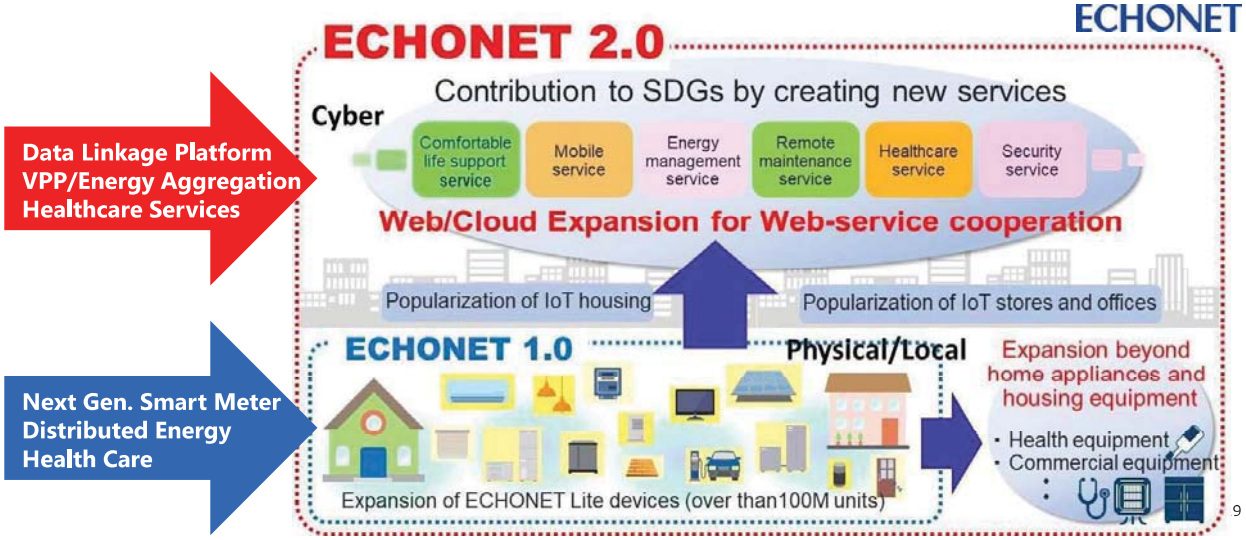
https://cuusoo.com/projects/50369/challenges/result, より抜粋

In the prediction contests held by electric power companies, we won the highest prize consecutively

9

ECHONET 2.0

From Home Appliance Network to Web Service Platform On Next Generation Smart Meter Era.



9

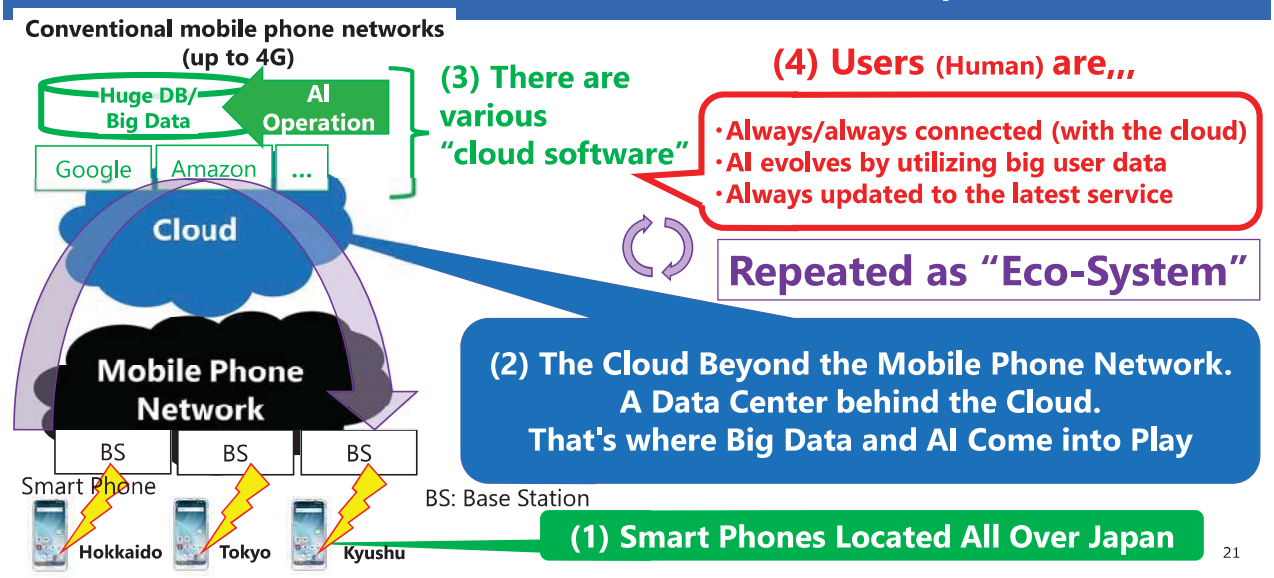
03 Software Defined Social Infrastructure

～Toward Edge Computing & Beyond-5G Era～

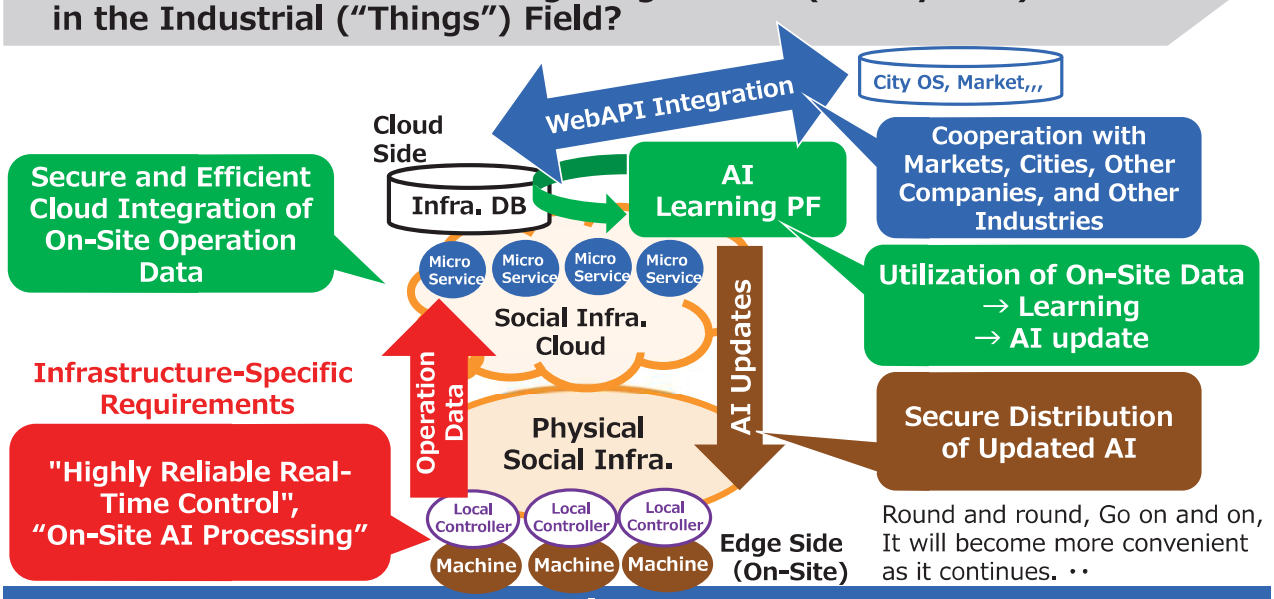
 20

Sources of Value in Big Data/AI: "Daily Operation" and "Learning and Growth"

Mechanism of the "human-based" ecosystem



How Can We Create a "Strategic Digitization (Eco-System)" in the Industrial ("Things") Field?



Strategic Edge/Cloud Collaboration and Its Data Distribution Infrastructure are Required

Expectations for Social Infrastructure x 5G/Local 5G

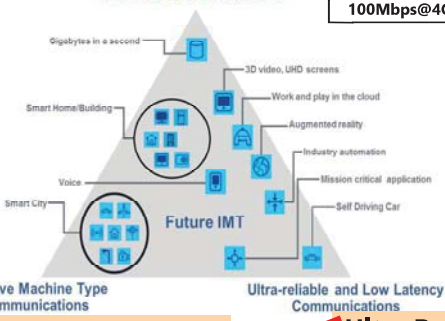
Faster, More Numerous, Less Latency, and More Reliable

More Than 4G

Super Fast(eMBB)
Peak Rate: 20Gbps
Effective Speed>100Mbps

Source:
IMT Vision：Usage Scenarios of IMT for 2020 and beyond

100Mbps@4G



New **Massive Communications (mMTC)** $10^6/km^2$ — 60k/km² @4G

New **Ultra Reliable Low Latency (URLLC)** <1ms — 5ms@4G

Moreover.,,,

Assign Occupied Frequency "Just for You"

No interference
→ High Reliability due to Dedicated Frequency

 23

Social Infra. ×5G: Bringing the Cloud Environment much closer to the "Edge"

MEC: Multi-Access Edge Computing

"Edge Cloud" Provided. AI can be Processed on the MEC

Conventional mobile phone networks (up to 4G)

Huge DB/ Big Data — AI Operation — Deep in the cloud

Google | Amazon | ···

Cloud

Mobile Phone Network

BS

Smart Phone

Local

Place "Edge Cloud" in Local

5G Era!

Central Cloud

Global DB/ Global Big Data — Global AI Operation

Google | Amazon | ···

Spreading Cloud

5G/L5G is All-in-One set

Local DB Local Big Data — Local AI Operation — Edge Cloud ←MEC

Google | Amazon | ···

BS

IoT

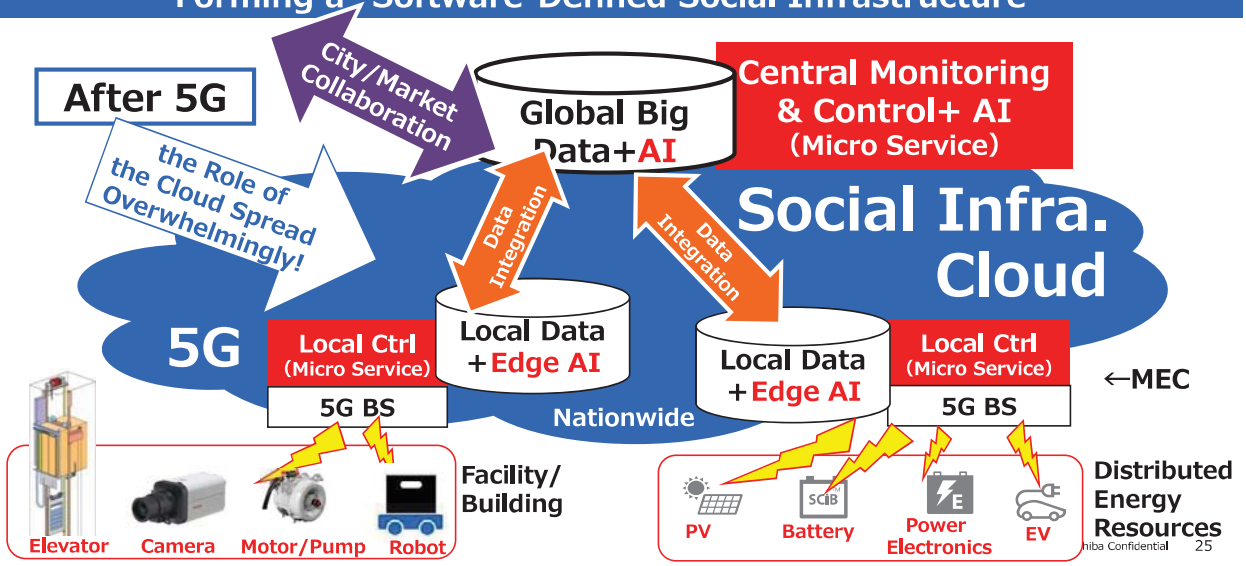
Smart Phone

High Reliability, High Speed, Real Time Wireless Communication

Social Infrastructure on Beyond 5G Era

MEC : Multi-Access Edge Computing

Cloud and 5G are Integrated to CPS-based Social Infrastructure Forming a "Software-Defined Social Infrastructure"

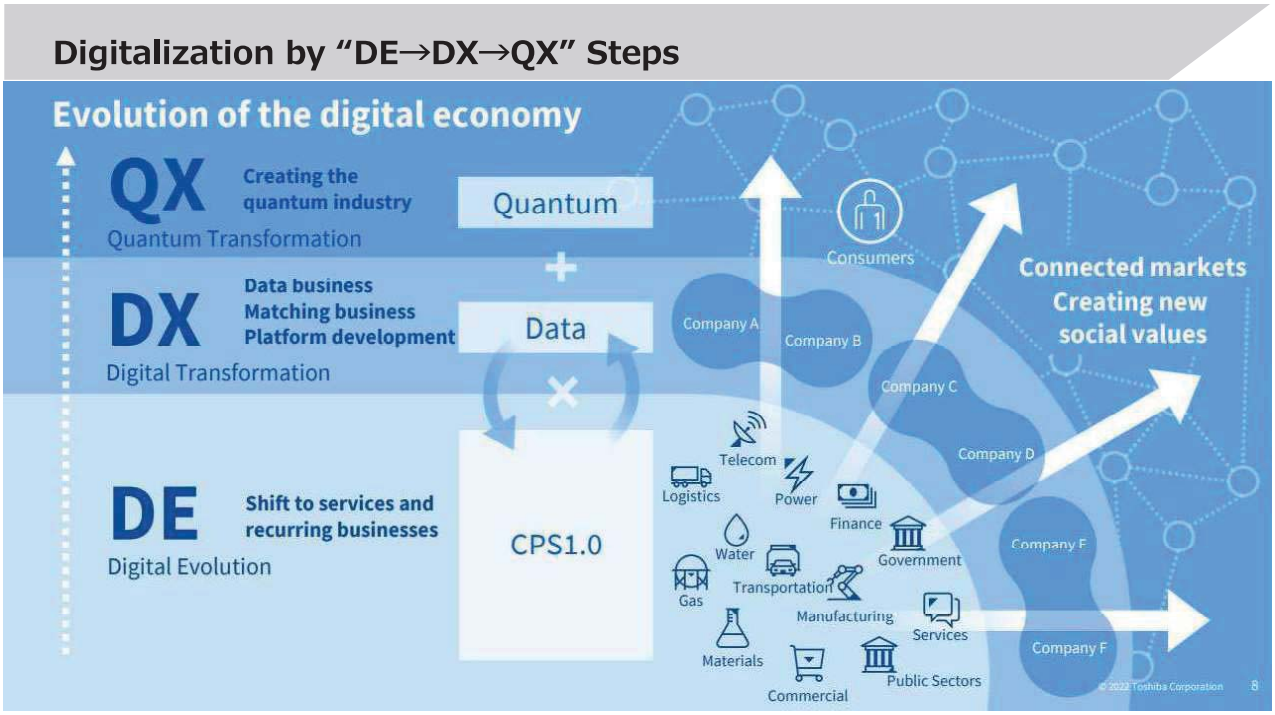


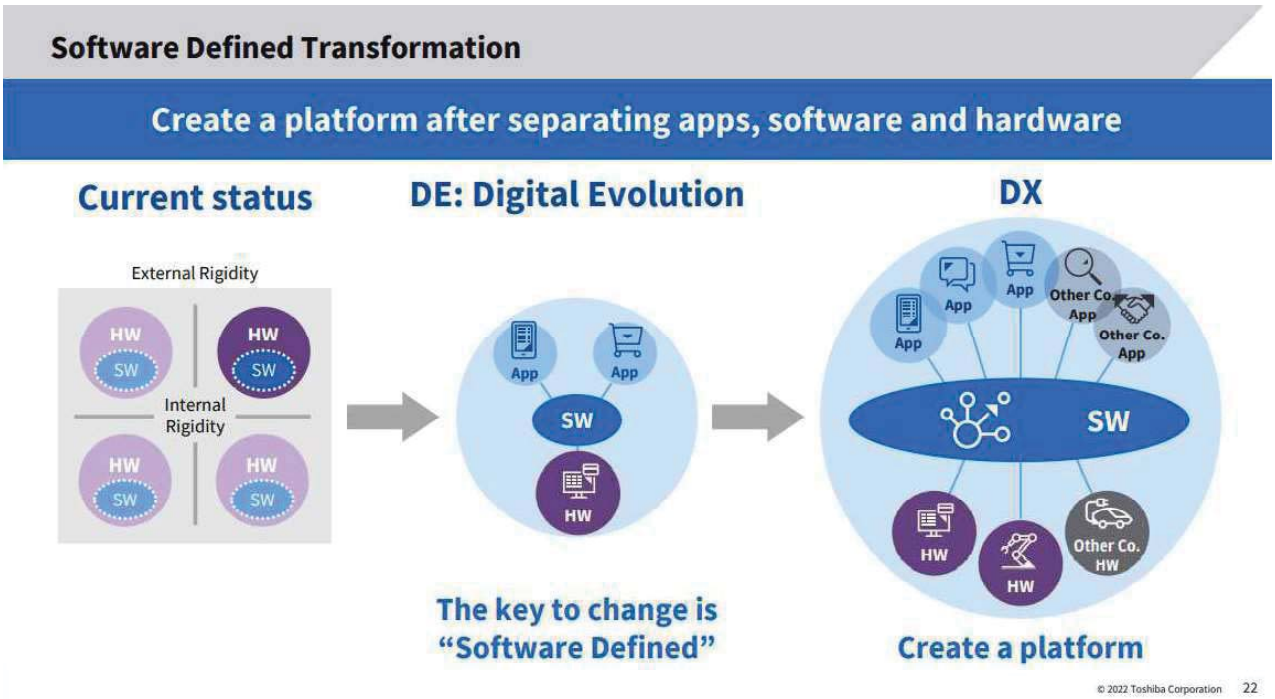
 25

04 Examples of Toshiba's Activity

～Toward Edge Computing & Beyond 5G Era～

 26


Digitalization by “DE→DX→QX” Steps
Evolution of the digital economy
QX
Creating the quantum industry
Quantum Transformation
Quantum
DX
Data business
Matching business
Platform development
Digital Transformation
Data
DE
Shift to services and recurring businesses
Digital Evolution
CPS1.0
Consumers
Connected markets
Creating new social values
Company A
Company B
Company C
Company D
Company E
Company F
Telecom
Logistics
Power
Finance
Water
Government
Gas
Transportation
Manufacturing
Services
Materials
Commercial
Public Sectors
© 2022 Toshiba Corporation
8


Software Defined Transformation
Create a platform after separating apps, software and hardware
Current status
DE: Digital Evolution
DX
External Rigidity
HW
SW
Internal Rigidity
App
SW
HW
Other Co.
The key to change is
“Software Defined”
Create a platform
© 2022 Toshiba Corporation
22

We shared many of our ideas at DICOMO2022 on July!

Thank you very much for great discussions there!

QX(Quantum Transformation) :

① QKD (Quantum Key Distribution)

② QSBM+ (Quantum inspired Computer on the cloud& FPGA)

DE(Digital Evolution) / DX(Digital Transformation) :

③WebAPI Test Automation for Distributed Energy Systems

④ FPGA implementation of AI algorithms (Image Recognition)

 29

More Ideas from Toshiba!

We show three more ideas from Beyond-5G viewpoints this time

(1) Software Defined Robot Concept based on MEC

(2) Microwave Wireless Power Transfer

(3) 5G/TSN integration Technology

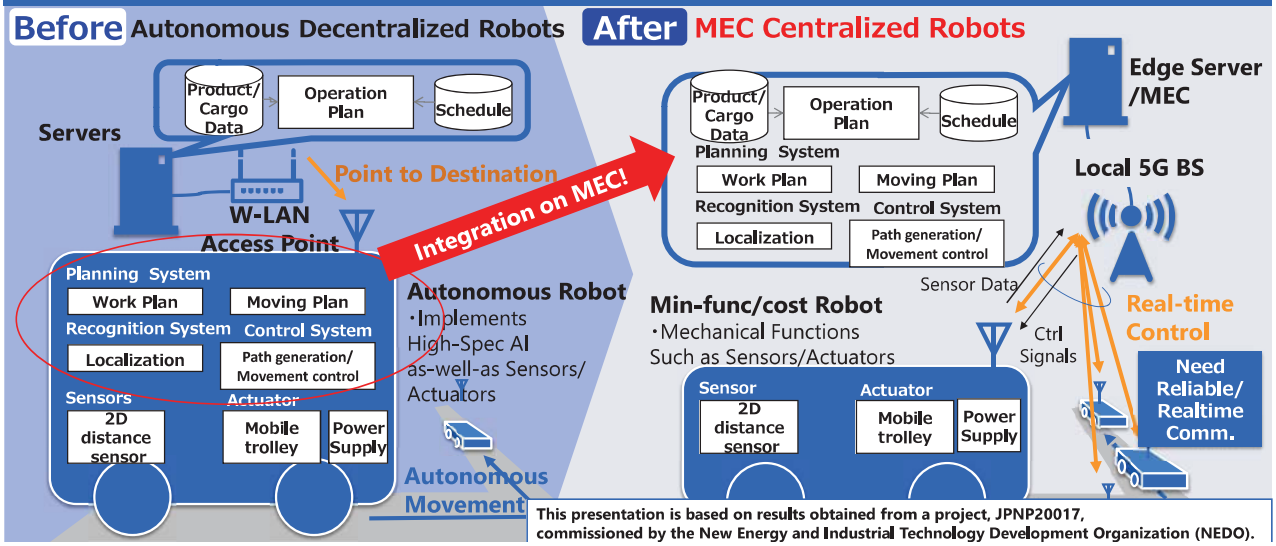
(Wireless=Wired=Virtual-Machine/MEC integration)

MEC: Multi-access Edge Computing
TSN: Time Sensitive Network

 30

(1) Software Defined Robot Concept based on MEC Tech.

Cloud/MEC Consolidation Aims to Reduce HW and Operation Costs and Improve Functionality such as Group Control/AI Learning



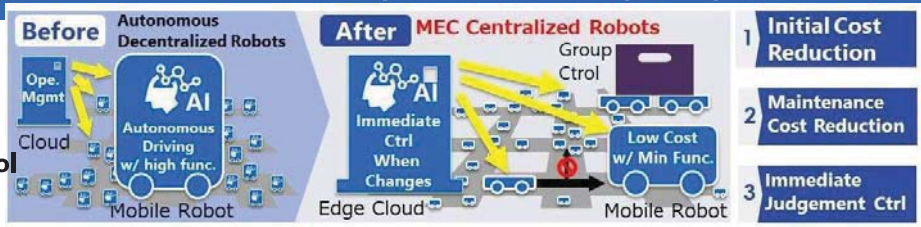
This presentation is based on results obtained from a project, JPNP20017, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

MEC: Multi-Access Edge Computing

(1) Software Defined Robot Concept based on MEC Tech

NEDO Research and Development Project of the Enhanced Infrastructures for Post-5G Information and Communication Systems/Feasibility Study

・**Aggregation of robot software/AI on the cloud/ MEC side**
・**Low-cost/ Real-time/ Simultaneous Group Control of Robot**



Local-5G enhanced

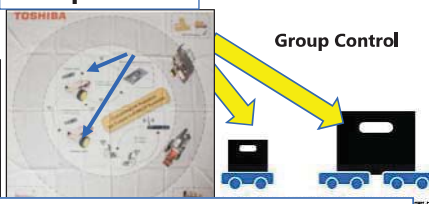
Local5G BS

Antennas

L-5G enhances Robot(AGVs)

High-Reliability Wireless

Real-time Group Control



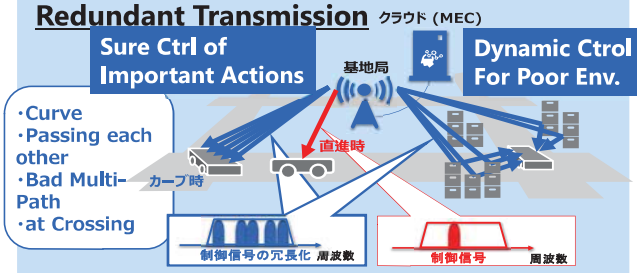
This presentation is based on results obtained from a project, JPNP20017, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

(1) High-Reliability Wireless Communication for SD Robot

Operation-linked/Environment-adaptive Redundant Transmission

Sure Ctrl of Important Actions

Dynamic Ctrol For Poor Env.



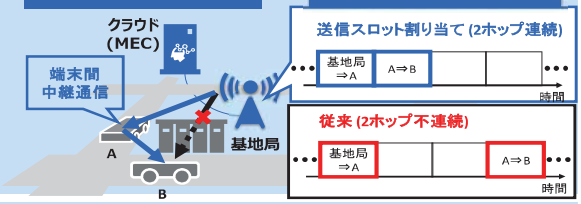
- Curve
- Passing each other
- Bad Multi-Path
- at Crossing

Relay Communication

Multi-hop Comm. of Ctrl Signals

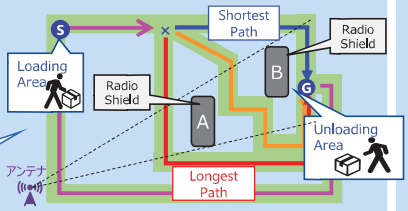
Sure Ctrol even on Dead Zone

Dynamic Slot Allocation



Dynamic Pathway Formation

Ever-Changing Radio-wave env. →Radio-Shielding Detection →Dynamic Change Route



Receive-Level Feedback & Dynamic Ctrl

Coordinated Transport

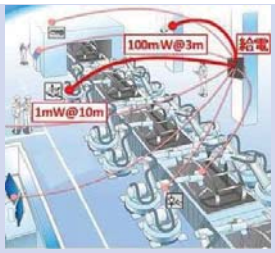
Coordinated transport for various sizes of parcels

Automatic Synchronization & Sure Control as single AGV

(2) Microwave Wireless Power Transfer

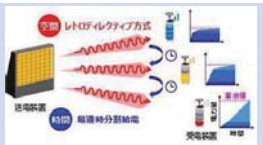
Wireless Everywhere! No need to change batteries!

Multi-Device/ Small-Size



Reduce Battery Change Cost

Optimization



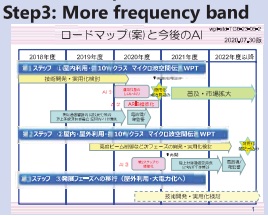
Intelligent Multi-Beamforming

Co-existence with legacy wireless

Legal Dev.

- Toshiba in charge of 5.7GHz
- Step1; Indoor, Manage people in and out

(Planned Step2: Outdoor/Manned
Step3: More frequency band



Prototyping

Multi-Sensor MWPT System

HD Camera Driving System

Even 1mW Sensors need battery-change every month

Passive Sensors

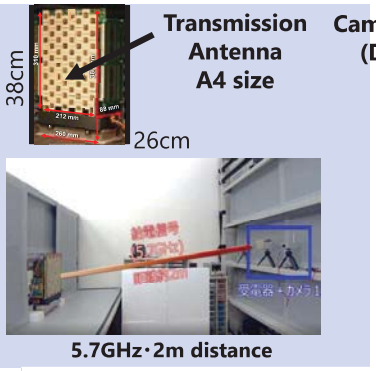
Active Sensors

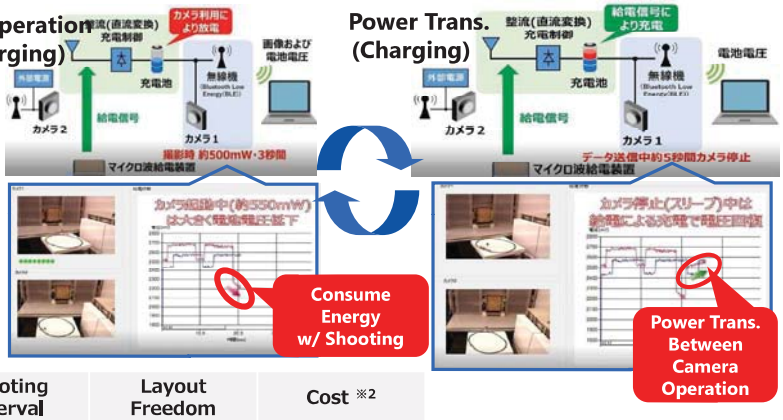
Active Wireless HD Camera

Cabinet Office : Cross-ministerial Strategic Innovation Promotion Program(SIP) “Energy systems of an Internet of Energy(IoE) society” 34

(2) Microwave Wireless Power Transfer (Prototype Implementation)

Drive HD Camera with MWPT





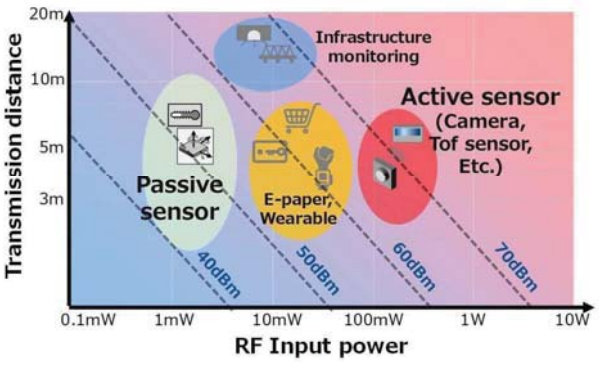
	Power Supply	Shooting Interval	Layout Freedom	Cost ※2
Toshiba	MWPT	A few sec(※1)	No Restrictions	○
A Company	Li aa Battery	1 hour※1	No Restrictions	◎
B Company	PV Panel	Depend on Env	Outdoor Only	△

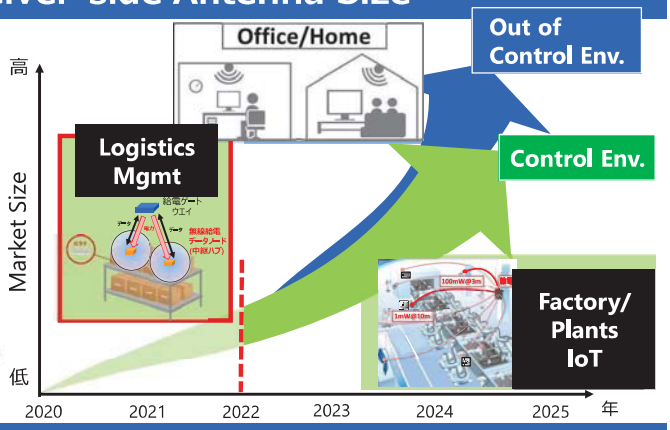
※1 Assume 2-years battery life.0.7s/1 shot, 2.8W

※2 Camera+Power Supply + Base Unit

(2) Various Applications and Future Prospects

Meet Various Power Level & Distance Requirement Depending on Receiver-side Antenna Size





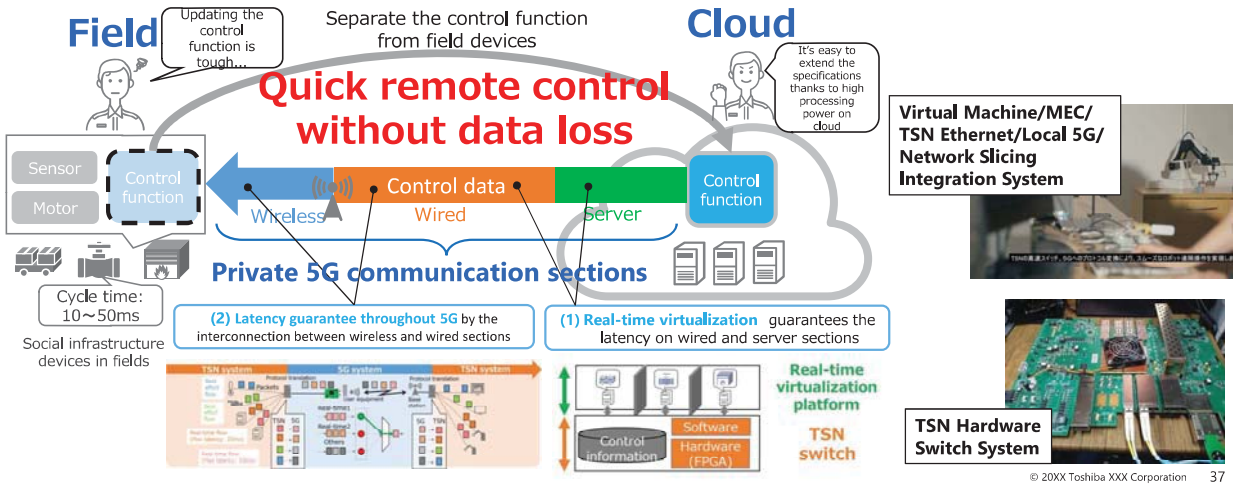
Step1: Secure Co-existence with Legacy Wireless & No Effect on Human Body Enhance to “Out of Control Environment” as Step 2/3.

(3) 5G and TSN Integration

TSN: Time-Sensitive Networking

Wireless=Wired=Virtual-Machine/MEC Real-Time Integration

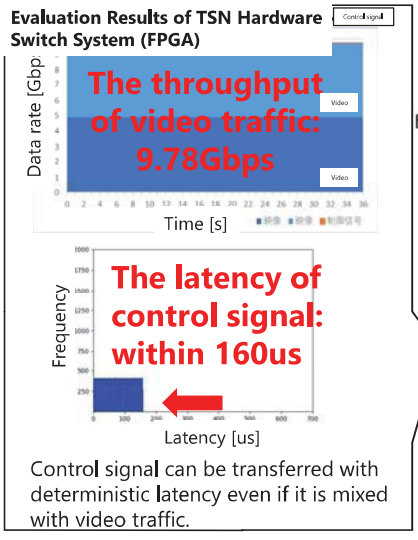
Real-time & High-Reliability Supervisory control system 5G enhanced Social Infrastructure



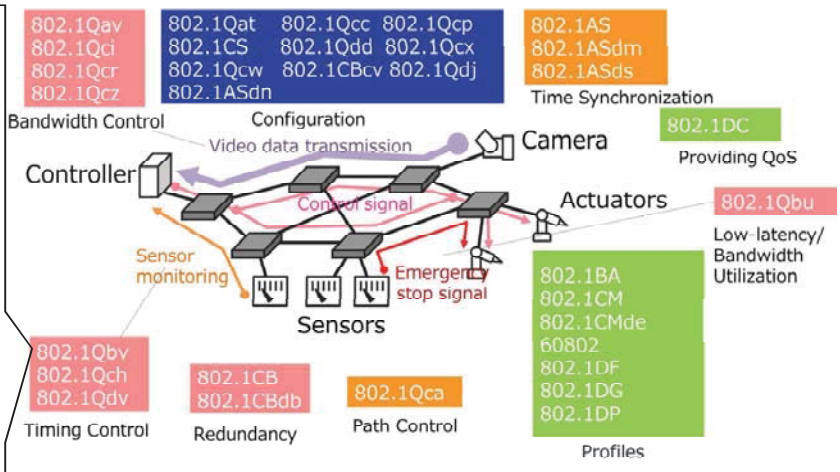
(3) 5G and TSN Integration: What is TSN?

IEEE 802.1 AVB/TSN: A set of standards that supports real-time and reliability on communication such as Ethernet.

Evaluation Results of TSN Hardware Switch System (FPGA)

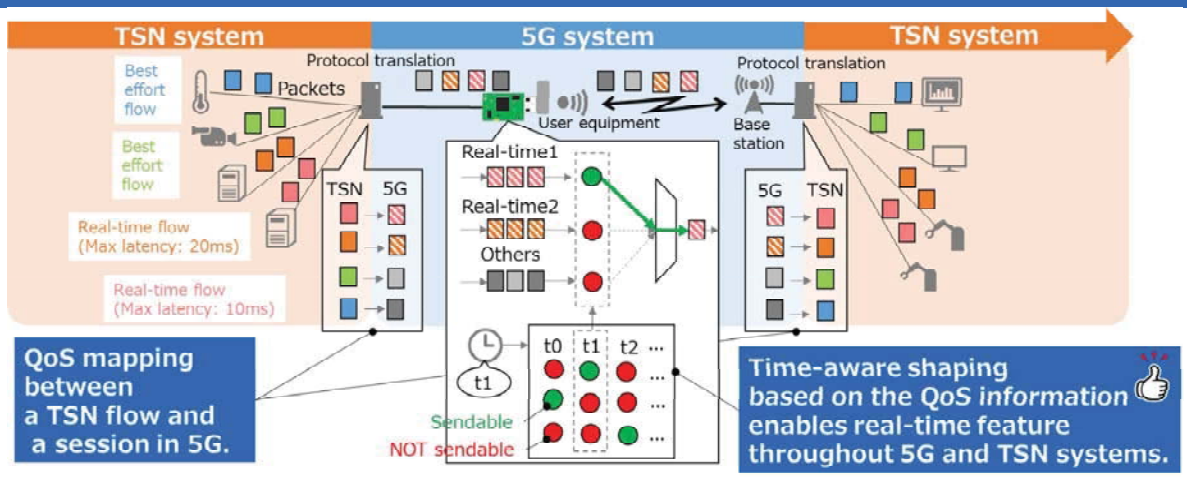


Control signal can be transferred with deterministic latency even if it is mixed with video traffic.



(3) 5G and TSN Integration

5G (Network Slicing & Signaling) & TSN protocol integration

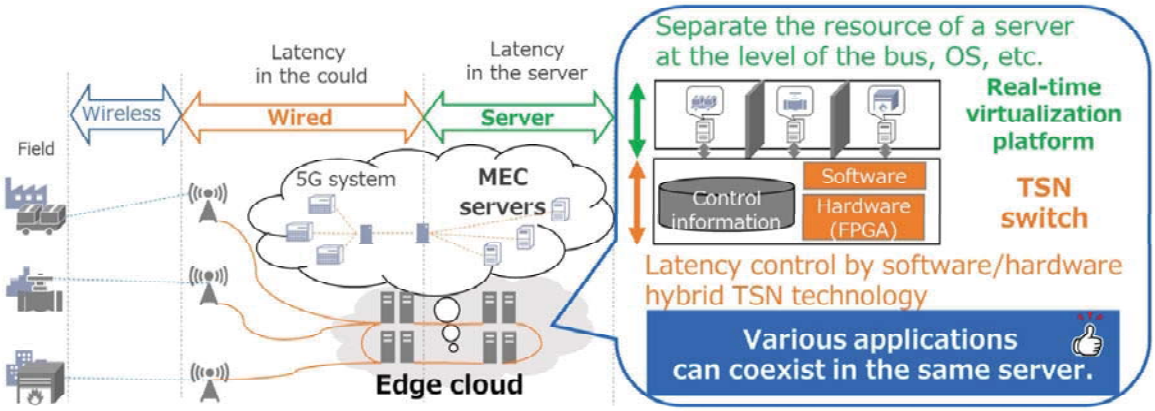


QoS: Quality of Service

 39

(3) 5G and TSN Integration

We also care inside virtual machine servers.
Guarantee End-to-End Real-time Communication.



- The TSN switch realizes deterministic communication in wired section.
- The real-time virtualization platform realizes deterministic communication in a server.

MEC: Multi-access Edge Computing

40

05 Committed to People, Committed to the Future.

 41

Toshiba Group's Vision

Committed to People,
Committed to the Future.

At Toshiba, we commit to raising the quality of life for people around the world, ensuring progress that is in harmony with our planet.

Future

For our children

Social Challenges: Achieving sustainability of people and the planet

Toshiba Initiatives: **Achieving carbon neutrality and a circular economy**

People

Safe, secure lifestyles for everyone

Poverty, human rights, disasters, disputes → **Building an infrastructure that everyone can enjoy**

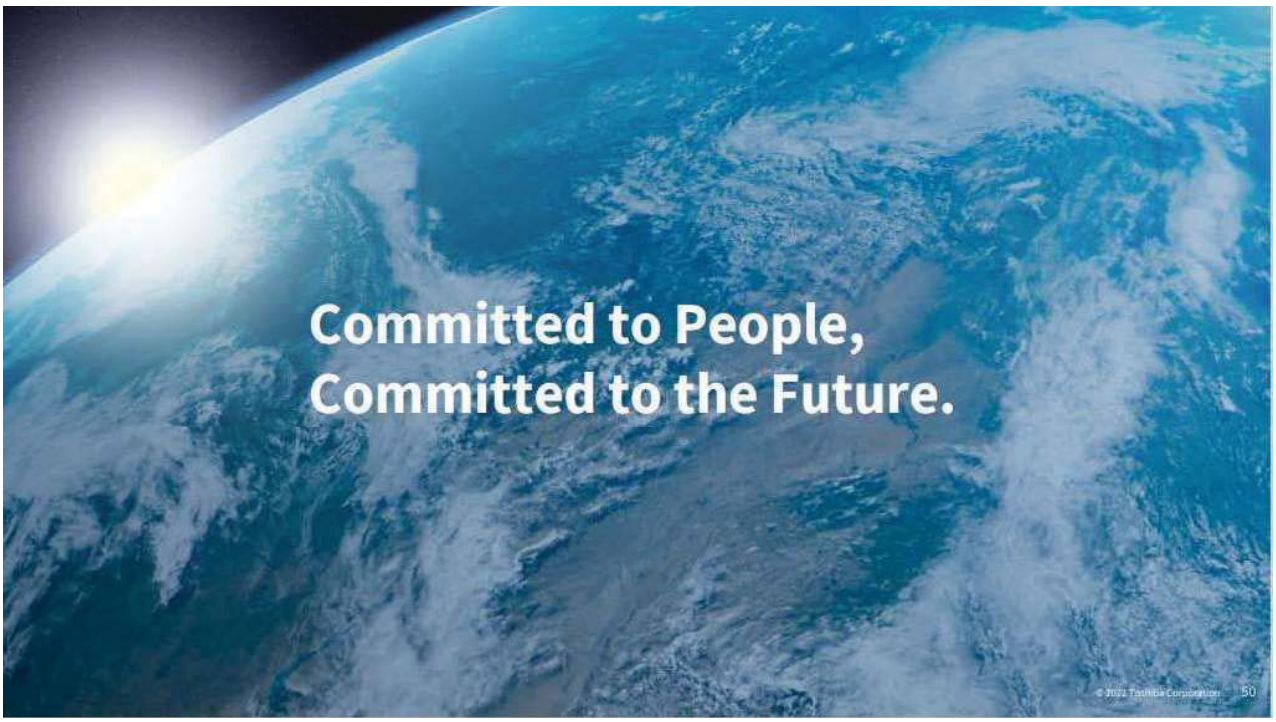
Planet

Social and environmental stability

Education, equality & fairness, climate change, resource depletion → **Building a society connected by data**

Contribute to the achievement of carbon neutrality & circular economy through digitization

 7


Committed to People,
Committed to the Future.
50

Session 7:
Application
(Chair: Tomoya Kitani)

Development of an Avatar Remote Communication System for ALS Patients Using Video Conferencing Tools

Nagisa Kokubu [*], Riku Sugimoto[*], Shouma Hamada[*],
Takayo Namba [**], and Keiichi Abe[*]
[*] Kanagawa Institute of Technology, Atsugi City, Kanagawa, Japan
[**] Kawasaki City College of Nursing, Kawasaki City, Kanagawa, Japan
abe@he.kanagawa-it.ac.jp

Abstract -Amyotrophic lateral sclerosis (ALS) patients with severe symptoms have difficulty in walking, going out, and talking with people. The authors believe that there is a need for a system that allows severe ALS patients to maintain contact with society and communicate with many people while staying at home. A representative previous study for solving this problem is an avatar robot. This avatar robot is placed in a place where the patient cannot go, and the patient can communicate through the avatar robot remotely by gestures. However, the avatar robot needs to be set up at the place where there is the person whom the patient is going to talk with, and the avatar robot needs to be removed after the conversation is over. Therefore, we propose a system that enables ALS patients with severe symptoms who have difficulty in going out to talk with many people via an avatar character easily and remotely for himself/herself. We have developed and evaluated a prototype of our proposed system.

Keywords: Remote Communication, ALS, online conferencing tool.

1. INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a disease in which the nerves that control voluntary movements (motor neurons) are damaged, resulting in the loss of muscles throughout the body and those necessary for conversation. In severe cases, the patient becomes bedridden and isolated, with no contact with other people, and it is difficult for the patient to move his/her body, although he/she is conscious. A previous study aiming to solve this problem is an avatar robot [1]. This avatar robot is placed in a place where ALS patients cannot go, and it can communicate with a person remotely by gestures. However, it is necessary to set up the avatar robot and remove it after the conversation. Therefore, the system requires help from other people besides the ALS patient.

In this study, we propose a system that enables ALS patients to communicate with many people in an enjoyable way, so that they can keep in touch with the society and communicate with various people. We have developed a system that enables ALS patients to talk with many people by their own will by using an eye control technology and an online conferencing tool[2]. Furthermore, the system allows ALS patients to choose their favorite avatar character during the conversation. We developed and evaluated a prototype of the proposed system. In order to evaluate the effectiveness of the proposed system, we conducted a questionnaire evaluation with an ALS patient.

2. RELARED WORKS

When ALS patients become severely ill, it becomes difficult for them to move their bodies and to speak. Conversation aids are an existing technology to help severely ill ALS patients communicate their intentions. The first representative product of portable conversational aids is a keyboard input type conversational aid [3]. This keyboard-input type conversation aid uses the keyboard to create sentences, which are then spoken by a voice synthesizer. It also has functions for registering frequently used words in categories and communicating using images. However, this keyboard input type conversation aid is difficult to use for severely ill ALS patients who have quadriplegia and have difficulty moving their bodies because the device requires the user to input data by hand. On the other hand, there is a gaze-input conversation aid [4]. Gaze input type conversation aids are devices that use a gaze input instead of a keyboard to create sentences and have them spoken by speech synthesis. Gaze-input conversation aids use a special gaze-input interface which is expensive and increases the system introduction cost. Common to these two products is the issue of not being able to converse with many people at once because they are face-to-face conversational types. As a previous study to solve these problems, there is an avatar robot developed by Oly Research Institute. This robot is installed in a place where ALS patients cannot go, and it can talk with a person whom the patient wants to talk with remotely by gestures. However, the robot needs to be set up at the place where there is the person whom the ALS patient wants to talk to, and the robot needs to be removed after the conversation is over, which requires the help of others besides the ALS patient. Another avatar technology other than robots is called Metaverse [5]. The avatars used in the metaverse are computer graphics, and the user can choose various characters of his/her choice. With this technology, people who have difficulty communicating in the real world can freely enjoy conversations and other activities in a virtual space through the avatar of their choice. In this metaverse, there is no need to set up an avatar like an avatar robot in advance at the place where you want to have a conversation, and no need to remove the avatar after the conversation is over. However, the interface used in the metaverse is currently a technology for able-bodied people, and it is difficult for people with physical disabilities, such as ALS patients, to use it because they cannot speak.

Therefore, this study proposes a system that solves these issues.

3. PROPOSED SYSTEM

Chapter 3 provides a detailed overview of the avatar remote communication system using video conferencing tools proposed in this paper.

3.1 Outline of the Proposed System

The proposed system is intended for ALS patients with severe symptoms who have difficulty in going out and enables them to easily and enjoyably talk with many people remotely without requiring help from others. The proposed system uses an online conferencing tool [2], as a remote communication tool for the ALS patient and the other party to have a conversation. The first reason for using an online conferencing tool is that it is widely used around the world as a video conferencing tool and is compatible with various platforms. The second reason is that ALS patients can easily communicate with other patients at home or in the hospital, regardless of their location, as long as they own a terminal that can use the online conferencing tool and have access to a network. Next, the proposed system uses an eye tracking interface [6] for the ALS patient's side. We chose the Eye Tracking interface because it is less physically demanding than the keyboard operation, since severely ill ALS patients have significantly reduced physical capabilities. The text input from the Eye Tracking interface was used as the text-to-speech interface. The system also allows the user to add and select a character for the voice dialogue agent that will be the avatar of the ALS patient by utilizing the function of adding and selecting voice dialogue agents in a previous study[7]. The ALS patient's favorite character can be added to the system by adding a GIF (Graphics Interchange Format) file of an animated character using FaceRig[8], etc. The system can also output different voices for each character by adding a voice model for each character and configuring the voice, speed of speech, and endings. The system allows the user to remotely communicate with the other party via the voice dialogue agent by transmitting the display screen of the voice dialogue agent selected by the user on the shared screen of the online conferencing tool. We call this proposed system Avatar Remote Communication System for ALS Patients (ARCS-ALS).

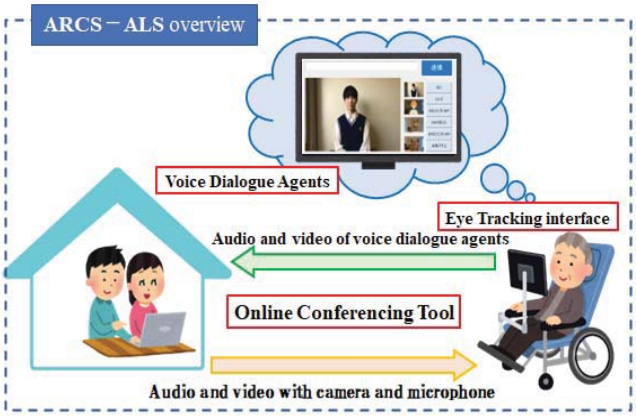


Figure 1: ARCS-ALS overview.

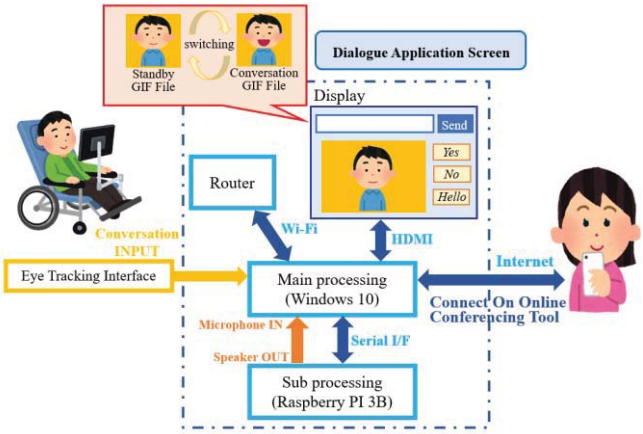


Figure 2: ARCS-ALS Prototype System Overview.

4. PROTOTYPE DEVELOPMENT

Chapter 4 details the development of the ARCS-ALS prototype system proposed in Chapter 3.

4.1 Outline of prototype system

Figure 2 shows an overview of the prototype system. In this prototype, distributed processing is performed by a Windows PC and a Raspberry PI3B+ (RPI). The Windows PC is in charge of the main processing and the RPI is in charge of the sub-processing. The main processing on the Windows PC includes the input processing of the Eye Tracking interface (EyeTracke4C,Tobii)[6] used by the ALS patient to operate the ARCS-ALS, the switching processing of the video for the voice dialogue agent that speaks on behalf of the ALS patient, and the processing of the video for the waiting and conversation. The sub-processing on the RPI includes switching the video between the standby and conversational mode, and remote communication with the conversational partner via the online conferencing tool. The RPI also outputs speech synthesized content input from the Windows PC in accordance with the voice of the voice dialogue agent and instructs the Windows PC when to switch the live-action video between the standby and conversational mode. The speech synthesis output of the sub-processing (RPI) is input to the microphone on the Windows PC side of the main processing via the speaker output of the RPI. This results in the synthesized voice being output to the other side via the online conferencing tool. Open J Talk[9] was used for the speech synthesis of the voice dialogue agent.

4.2 Application screen

In the prototype of the proposed system, ARCS-ALS, we developed a voice dialogue application to be used by ALS patients when conversing with each other. Figure 3 is a general view of the prototype system developed for this project, and Figure 4 shows the screen of the application developed this time. As shown in Figure 4, the large character in the middle of the screen is the selected character. In other words, it will be a voice dialogue agent that will carry out conversations on behalf of the ALS

patient. When the patient wants to change the selected agent to suit his/her preference, he/she can click on the character display on the right side of the screen to change the character automatically.

In this prototype, we prepared four characters: the user's own image, a child character, and animal characters of a dog and a cat. Two methods were prepared for the input of the conversation. In the first method, frequently used conversations such as replies, and greetings are registered in advance and can be spoken at the touch of a button. The second method is to use the keyboard, which is one of the functions of the eye control input interface, to input text. In this method, the ALS patient inputs what he/she wants to say using the keyboard and presses the "send" button to speak. Although it takes some time to input the content of the conversation, any kind of conversation can be sent.

4.3 Eye tracking interface

ARCS-ALS uses the Eye Tracking interface[6] as a method for ALS patients to talk. The reason for using this interface is that even ALS patients who have limb disabilities and have difficulty moving their bodies can use the Eye Tracking interface for eye control if they do not have eye problems. In our proposed system, we used an off-the-shelf eye tracking interface and the eye tracking function[3] that comes standard in Windows 10. To use the system, the Eye Tracking interface is connected to a Windows PC via a USB interface, and the main unit of the Eye Tracking interface is attached to the bottom of the monitor. After installing the driver, adjusting the position of the Eye Tracker 4C, and performing simple setups such as a visibility test, the system is able to detect the user's line of sight on the Windows PC screen. Next, by activating the eye control function that comes standard with Windows 10 and later, the user can click, use the mouse, and perform keyboard input, so even the physically challenged can operate the PC using only their line of sight. The Eye Tracking interface should be installed at an appropriate distance from the user, which is approximately 90 cm. If the distance is too close or too far, the camera cannot be calibrated accurately and may not operate properly. Figure 5 shows the eye tracking interface used in the prototype and the eye control function. The eye tracking interface attached to the lower part of the monitor (circled in red) recognizes the user's line of sight, and the eye control function (circled in green) is used to control the operation. By looking at the

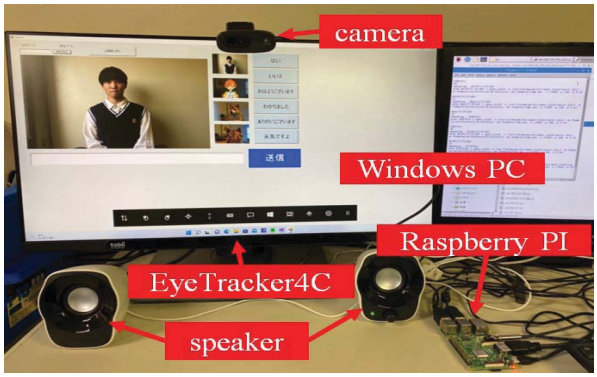


Figure 3: Prototype system developed for this project.

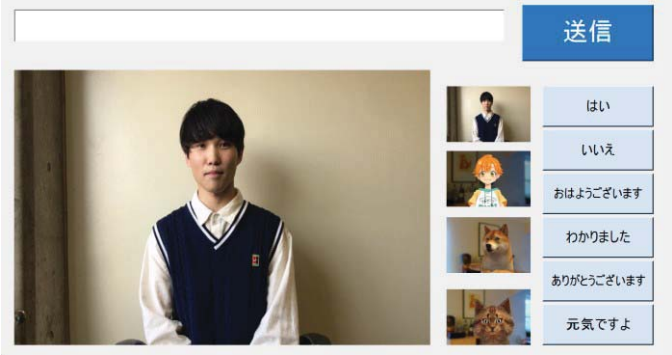


Figure 4: Application screen for ARCS-ALS.

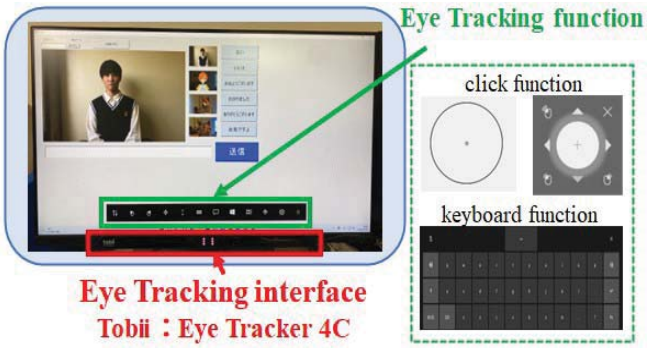


Figure 5: Overview of Eye Tracking interface.

desired function for a certain period of time, the user can instruct the eye tracking function to display icons to select the desired location for clicking and a keyboard for inputting text, as shown in Figure 5. At first, it may be a little difficult to operate the function you want to use, but as you gradually become accustomed to it, you will be able to operate it at will.

4.4 Voice dialogue agent

In the ARCS-ALS system, a voice dialogue agent is used which talks to the other party instead of the patient. The voice dialogue agent was created by applying our previous studies. In this prototype, we prepared four characters: a character created from the ALS patient's own image, a child character, and animal characters (a dog and a cat). This allows the user to choose one of the characters according to his or her preference and environment, such as those who want to show their healthy self when talking to other persons or those who want to appear to be a substitute cartoon character instead of their own image. The voice dialogue agent can be created in two main ways: as a CG character or as a live-action character. Figure 6 shows a summary of each method. First, CG characters were created using FaceRig[8] software, which allows the user to become any character using a webcam, and two GIF images were created: one in the standby mode (as if the user is listening to a conversation) and one in the conversational mode (as if the user is talking). By displaying the two GIF images in accordance with the speech state of the voice dialogue agent using text-to-speech output and switching the character, the mouth movements are synchronized with the video as if the character is actually speaking. The live-action character is an application of our previous work on creating CG

characters. To create a live-action character, two live-action videos were actually shot using video equipment, one in the standby mode and the other in the conversational mode. As in the previous study, a live-action character is easily created by switching between the two GIF images in accordance with the speech state. In this case, it is necessary to shoot a short movie of 3 to 5 seconds and change the movie file to a GIF file. This makes it possible to repeat and play back the short video, which allows for longer text conversations. This method has the advantage that, once the user becomes accustomed to it, a live-action character can be created in about 10 minutes, making it possible to create a character of the person to be reproduced at a low cost. In addition, it is considered to be better to create the image of the ALS patient before the disease becomes severe, because it is considered to be practically difficult for the ALS patient to film himself after the disease has become severe.

4.5 Voice synthesis output

The voice of the voice dialogue agent, which is the avatar of the ALS patient, is synthesized by Open J Talk [9], and its output is shown in Figure 7. In this system, the ALS patient uses the eye control input interface to input the content of the conversation in text using the application developed in this study. The input conversation content is output as text-to-speech by Open J Talk [7], and the voice dialogue agent speaks the content. Acoustic models are prepared for each of the voice dialogue agents selected by the user, and the voice, speech style, and endings are changed for each character to speak. The dog CG character would end with " bow-wow" and the cat CG character would say " meow". On the other hand, when using a live-action character created from the ALS patient's own video, we prepared an acoustic model that closely resembled the patient's own voice, and then manually adjusted the parameters using Open J Talk to make the voice sound similar to the patient's own voice. Specifically, the voice quality, which can be changed to feminine or masculine by changing related parameter values, the pitch shift, which changes the tone of the voice, and the speech speed, which changes the speed of speech, were adjusted to match the patient's voice. The values of these voice parameters were used because they can be fine-tuned, which actually takes time but produces highly reproducible voice. Other speech synthesis technologies include those that utilize AI technology to learn voices. Coestation [10] and LYREBIRD [11] are representative of these technologies. With these technologies, it is easy to produce a voice synthesizer that is close to the user's own voice by reading a few examples of sentences and having the AI learn those sentences. However, since the example sentences need to be learned repeatedly, it is not possible to train the AI for ALS patients who have difficulty in speech. Therefore, this system uses Open J Talk, which allows the user to manually adjust the parameters of the acoustic model based on the original voice.

5. EVALUATION

In this chapter 5, a questionnaire evaluation was conducted to evaluate the prototype of the proposed system.

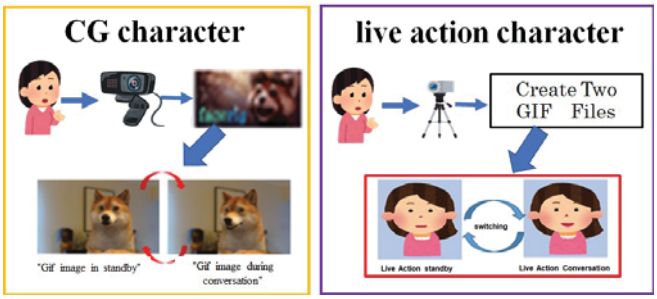


Figure 6 : How to Create Voice Dialogue Agents.

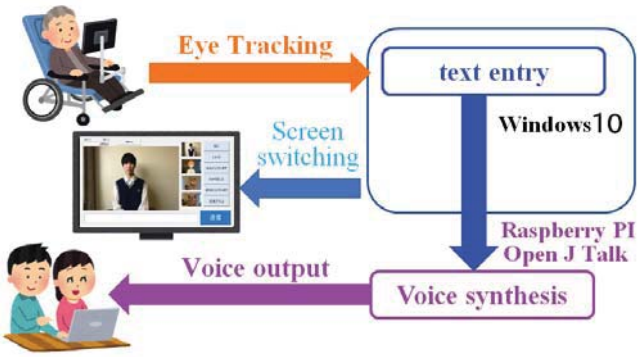


Figure 7: Speech synthesis output mechanism.

This time, the evaluation was conducted the questionnaire evaluation for ALS patients. From the questionnaire evaluation, we investigated the effectiveness of the proposed ARCS-ALS system, the functions required for the system, and points to be improved, and we describe the details.

5.1 Method of conducting questionnaire evaluation

To ask for an evaluation of the proposed system, we contacted one ALS patient through an acquaintance who has a close relationship with ALS patients. Because of the ongoing COVID-19 situation and the difficulty of face-to-face evaluation, we asked the patient to watch a video (about 9 minutes) summarizing the use of the prototype of the proposed system and then to answer the questionnaire.

5.2 Evaluation Result of ALS patients

Table 1 shows the evaluation results. In response to the question 1, "Would you like to use the system shown in the video?", the answer was "Yes". In response to question 2, "Would you like to use the system you saw in the video?", the answers were "I can use it by myself," "I can communicate easily," "I can communicate with others as my favorite character," and "I can talk with many people even if I can't go out". In addition, "It is easy to listen to because it is speech synthesized and the pronunciation can be inflected" and "The predictive conversion seems easy to use". Next, to question 3, "What would you like to use as a character for the spoken dialogue agent in this system?", the respondent answered, "My current live-action movie" ,"My favorite animal character (dog, cat, etc.)", "My favorite animation", and so on. To the question 4, "What would you like to use as the voice of this system?", the answers were "my voice", "favorite star's voice", "favorite animation", and so on. There were no responses to the question 5, "What

improvements do you think are needed in the layout of the application screen of the system?". To the question 6, "In addition to the remote conversation function, what other service functions would you like to see in this system?", other comments included "a doorbell and zoom function," and "more types of keyboards". Next, there were no responses to the question 7, "What do you think needs to be improved in the future?" Finally, to the question 8, "Please feel free to give us any other comments or requests you may have", the answer was "I use the Internet, YouTube, Facebook, LINE, and recently Google Sheets instead of Excel. I rarely use other apps, so I thought it would be better to create a well-balanced system. Please let me reserve my judgment until I actually use the system.

6. CONSIDERATION

From the results of the questionnaire evaluation of the ALS patient, in response to the question 1, "Would you like to use the system you saw in the video?", the answer was "Yes" . This indicates that the system is probably effective even from the viewpoint of ALS patients. Next, from the results of Question 2, we can expect that the system can be used in various situations, since the respondent selected all the items as the reasons why the patient would like to use the system. Next, from the results of Question 3, the respondent selected all the options for the character of the spoken dialogue agent in this system except for "live-action videos of my past when I was healthy". This may suggest that ALS patients do not want to return to what they were before they became ill. Next, from question 4, the respondent selected all of the items the respondent wanted to use as the voice of the system, which suggests that the respondent want the voice to match the character. Next, from Question 5, the layout of the application screen of the system was evaluated to be good, and there was no point of improvement from the viewpoint of the ALS patient. Next, from the question 6, what kind of service functions other than the remote conversation function should be added to this system, the respondent selected all of the options, which confirms that more functions are required. In particular, it is considered necessary to add functions that make life more convenient, such as remote control of home appliances and health management functions. The result of Q7 suggests that the respondent found none of the listed items needed to be specifically mentioned as an item that should be further improved. Finally, the results of Question 8 showed that ALS patients use various tools such as the Internet, YouTube, and SNS. From this result, we can assume that the ALS patient who responded to this questionnaire has considerable knowledge about computers. We would like to increase the number of subjects in the future, since we had only one subject this time.

7. FUTURE WORK

From the evaluation of the ALS patient's questionnaires, we were able to understand what improvements are necessary for the proposed system. Based on the results of the evaluation, we would like to make improvements toward the practical application of the proposed system. We had only one ALS patient as the respondent in the questionnaire evaluation this time, so we would like to increase the number of questionnaires in the future. In addition, we would like to evaluate the proposed system from multiple viewpoints by conducting questionnaire evaluation for doctors who are engaged in ALS treatment.

Table1:Results of Questionnaire Evaluation of ALS Patient.

question	choice	Answer
Q1: Would you like to actually use the system you have seen in this video? (Or would you like to use the system you saw in the video?)	1. yes 2. no	1
Q2. If you answered "yes" to Q1. In what ways would you like to use it?	1. because I can use it by myself 2. because it is easy to communicate 3. because I can communicate with others as my favorite character 4. because I can communicate with many people even if I cannot go out 5. other	1 2 3 4 5 [It is easy to listen to the voice synthesizer, and the pronunciation can be inflected. Predictive conversion is also easy to use.]
Q3. What would you like to use as a character for the spoken dialogue agent in this system? Please select more than one.	1. past live-action videos of yourself when you were healthy 2. live-action video of yourself in the present 3. favorite animal (dog, cat, etc.) character 4. favorite cartoon or other character 5. other	2 3 4 5 [Furniture, musical instruments, and food]
Q4. what would you like to use as the voice of this system? Please select more than one.	1. your own voice 2. voice of your favorite star 3. voice of your favorite cartoon character 4. other	1 2 3
Q5. What improvements do you think are needed in the layout of the application screen of the system? Please select more than one that apply.	1. difficult to select canned text 2. the text box is too small and it is hard to input characters 3. difficult to select characters 4. small screen size 5. small character size 6. other	
Q6. In addition to the remote conversation function, what other service functions would you like to see in this system? Please select more than one that apply.	1. subtitling of conversations 2. function to remotely control home appliances 3. function to measure and manage the user's health (body temperature, pulse wave, Spo2, etc.) 4. function to provide daily life information such as weather, time, etc. 5. other 5. other	1 2 3 4 5 [More doorbells, zoom function, and keyboard types]
Q7. What do you think needs to be improved in the future? Please select more than one.	1. speed of conversation 2. awkwardness of response 3. gaze input 4. application 5. screen size 6. audio quality 7. other	
Q8. Please feel free to give us any other comments or requests.	I think it is better to start with less, as more features will cost more. I use the internet, YouTube, Facebook, LINE, and recently google spreadsheets instead of Excel. I rarely use the rest, so I thought it would be better to create a well-balanced system. Please let me judge the rest by actually using it.	

ACKNOWLEDGEMENTS

This study was approved by the Ethical Review Board for the use of human subjects of Kanagawa Institute of Technology (No.20220715-01).

REFERENCES

[1] "Orihime Product Introduction ", < https://orihime.orylab.com/>, (Referred May.2022).

[2] “Zoom”, < https://zoom.us/jp-jp/meetings.html>, (Referred May.2022).
[3] “TALKING AID+”, <https://www.talkingaid.net/products/ta-plus>, (Referred May.2022).
[4] "LUCY Double Giken ", <http://www.j-d.co.jp/fukushikiki-lucy.html> , (Referred May.2022) .
[5] "Metaverse Japan", <https://metaverse-japan.org/>, (Referred May.2022).
[6] Eye Tracker Product List|Tobii Technology https://www.tobiipro.com/ja/product-listing>,(Referred May.2022).
[7] Nagisa Kokubu Tetsuto Mukai, Keiichi Abe, " A Proposal for a Low-Cost voice Dialogue System Using Live-Action Video Contents ", 2021 IEEE Global Conference on Consumer Electronics Proceedings (GCCE2021), pp.340-341, Oct. 2021.
[8] FaceRig Live2D Module, <https://www.live2d.com/interviews/facerig>, (Referred May.2022).
[9] Open J Talk, < http://open-jtalk.sp.nitech.ac.jp>, (Referred May.2022).
[10] coestation, <https://coestation.jp/consumer/coestation-napp/>, (Referred May.2022).
[11] Lyrebird, <https://www.descript.com/lyrebird>, (Referred May.2022).

Path Planning considering Pedestrian Characteristics

Yuto Yada*, Shunsuke Michita*, Seiji Komiya* and Toshihiro Wakita*

*Graduate School of Engineering, Kanagawa Institute of Technology, Japan
s2284001@cco.kanagawa-it.ac.jp

Abstract - Autonomous mobility in mixed traffic environments with pedestrians need functions to avoid contact with pedestrians. In this study, path planning method adapted to pedestrian face direction was developed. For pedestrians who are aware of mobility (forward facing walking), small avoidance path is generated. For pedestrians who are unaware of mobility (downward facing walking), such as those who are walking on their smartphones, large avoidance path is generated. Subjective evaluation experiments were conducted on four items: distance, speed, smoothness of avoidance, and reliability. The subjective evaluation results showed that the evaluations improved for all items except speed, both for forward and downward walking. In particular, for downward facing pedestrians the evaluation of the distance was considerably improved.

Keywords: Autonomous Mobility, Pedestrian behavior prediction, Yolo

1 INTRODUCTION

In recent years, the practical application of autonomous mobility has been progressing worldwide in areas such as office building security and package delivery. Autonomous mobility move in mixed traffic environments with pedestrians need a path planning function that avoids contact with pedestrians. When humans pass each other, they unconsciously make eye contact with each other and anticipate the other's movements to ensure smooth movement. Therefore, we are working on the realization of autonomous mobility that enables this type of behavior.

DWA (Dynamic Window Approach) [1] and RRT (Rapidly exploring random tree) [2] have been widely used as static obstacle avoidance methods for mobile mobility. The problem with these previous studies was that dynamic obstacles such as pedestrians could not be avoided because they were not considered.

For dynamic obstacle avoidance, pedestrian prediction using the Kalman filter[3], pedestrian prediction and avoidance using the potential method[4], and the application of ORCA (Optimal Reciprocal Collision Avoidance) to the prediction and avoidance of multiple pedestrians[5] have been studied.

One of the problem for previous studies is avoidance for pedestrians walking on their smartphones. It is difficult for mobility to avoid pedestrians walking while gazing at their smartphones. This is because their walking path is unstable and behavior prediction is difficult. Also, pedestrians may be surprised when mobility suddenly appears in their field of vision when they pass by at close range while they are gazing at their smartphones. To solve the problem, a method of warning by sound can be considered. Although pedestrians may notice mobility with sound warnings, this method causes mobility to impede pedestrians' walking, and frequent warning sounds can make pedestrians uncomfortable. Especially when autonomous mobility increases in the future, it is unlikely that autonomous mobility will always be prioritized over pedestrians. As another means, a method of avoiding large can be considered. Although the method could avoid the pedestrian safely, such large avoidance would be excessive for pedestrians walking forward. If excessive avoidance is always performed, there is a high possibility that it will take a long time to arrive at the destination or the route cannot be generated and the mobility cannot move.

Therefore, in this study, a method to adjust the amount of avoidance according to the face direction was attempted. After predicting the pedestrian's behavior, the risk of collision is reduced by avoiding a small amount when the pedestrian's face is in front of the vehicle and a large amount when the face is facing downwards. Despite avoiding pedestrians using this method, if a collision is unavoidable, the mobility stops. This avoidance strategy is similar to that used by humans every day.

2 METHOD AND EXPERIMENTAL EQUIPMENT

2.1 Method

To safely avoid a downward-facing pedestrian, the face direction of the pedestrian is recognized by face direction recognition. Next, if the result of the face direction is a downward-facing pedestrian, a large avoidance path is generated. We considered these two requirements for pedestrian avoidance.

The method is based on the following procedure.

Step1 Measurement of pedestrian position using 3D LiDAR, pedestrian detection, tracking, and pedestrian behavior prediction.

Step2 Pedestrian face direction detection by image recognition .

Step3 Collision risk area calculation based on the pedestrian behavior prediction and pedestrian face direction detection results.

Step4 Pedestrian avoidance route generation.

A schematic diagram of the proposed algorithm is shown below (Figure 1).

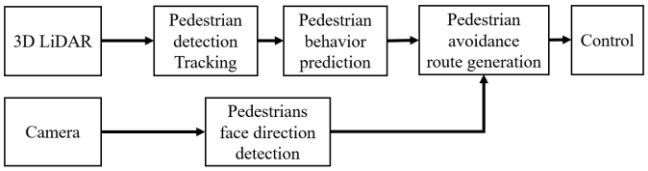


Figure 1: Method of pedestrian cooperative path planning

As in previous studies [3], 3D LiDAR information and a Kalman filter were used for pedestrian recognition and pedestrian behavior prediction.

2.2 System Configuration

The autonomous mobility used in this experiment is shown in Fig. 2. It was equipped with a camera for face direction detection of pedestrians and an omnidirectional laser sensor for self-localization and obstacle detection. Data obtained from these onboard devices is processed and controlled by a compact computer DH310 (Shuttle) and a Jetson Xavier NX (NVIDIA) (Table.1).

Figure 2: Mobility

Table 1: System Configuration

Camera	C920n web camera (Logicool)
LiDAR	VLP-16 (Velodyne)
Computers	DH310 (Shutlle)
	Jetson Xavier NX (NVIDIA)

2.3 Face Recognition

To estimate whether pedestrians are aware of autonomous mobility or not, this study assumes that pedestrians whose faces are forward-facing are aware of autonomous mobility and those whose faces are downward-facing are not aware. Pedestrian face direction recognition was performed using deep learning with Yolo [6]. An example of recognition is shown in Figure 3. First, 300 images were taken of each pedestrian with a forward face and a downward face. Next, 4000 epochs of deep learning by Yolo were performed using the collected images. The recognition rate of forwarding-facing was about 99% and that of downward-facing was about 77% (Table.2).

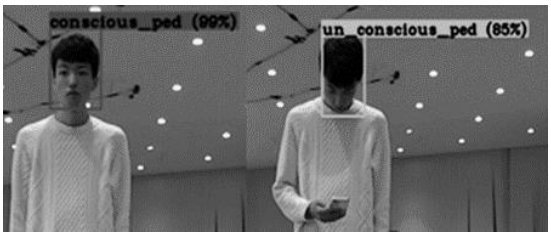


Figure 3: Recognized example of face direction

Table 2: Recognition result of face direction

		Recognition result	
		Forward	Downward
Human behavior	Forward	99%	1%
	Downward	23%	77%

2.4 Path Planning

Route generation was performed by RRT* [7] based on the occupancy grid map.

Based on the results of pedestrian face direction recognition and pedestrian behavior prediction, collision risk areas were defined on the occupancy grid map used in route generation (Fig. 4). For pedestrian behavior prediction, the walking speed of the tracked pedestrian was calculated using a Kalman filter, and the predicted position was calculated based on the calculated walking speed.

The width of the collision risk area is the same as the width of the body when avoiding a forward-facing pedestrian (hereinafter referred to as 'small avoidance'), and is wider when avoiding a downward-facing pedestrian (hereinafter referred to as 'large avoidance') so that a safe distance is maintained between the pedestrian and the collision risk area. The collision risk area was treated like an obstacle in the route generation.

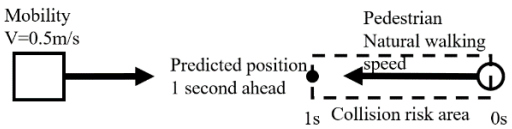


Figure 4: Collision risk area

3 EXPERIMENTAL RESULTS

Using the proposed path planning algorithm, a subjective evaluation experiment on pedestrian avoidance was conducted in a laboratory. A course was created as shown in Fig. 5, and the autonomous mobility was moved at a translational velocity of 0.5 m/s. Pedestrians were instructed to walk at their natural walking speed and evaluate whether the autonomous mobility could avoid pedestrians.

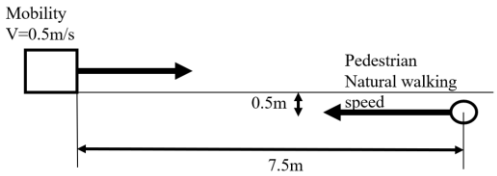


Figure 5: layout of pedestrian avoidance experiment

3.1 Path Planning Results

The path planning results are shown below. For comparison, similar experiments were conducted under path planning without behavior prediction. Three types of path planning were used: without behavior prediction (Fig. 6), small avoidance (Fig. 7), and large avoidance (Fig. 8). The bold lines are the selected paths and the branches are the candidate paths. It was confirmed that the system generated a largely avoidable path for downward-facing pedestrians compared to forward-facing pedestrians.

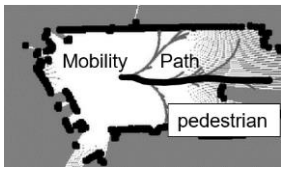


Figure 6: Path planning example of without behavior prediction path planning. The bold line represents the selected path and the branch line represent candidate path.

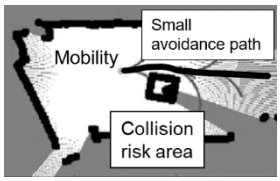


Figure 7: Path planning example of small avoidance cooperative path planning. The bold line represents the selected path and the branch lines represent candidate path. The square represents the collision risk area.

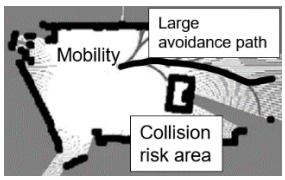


Figure 8: Path planning example of large avoidance cooperative path planning. The bold line represents the selected path and the branch lines represent candidate path. The square represents the collision risk area.

Figs 9, 10, and 11 show the trajectory examples for without behavior prediction, small avoidance, and large avoidance, respectively.

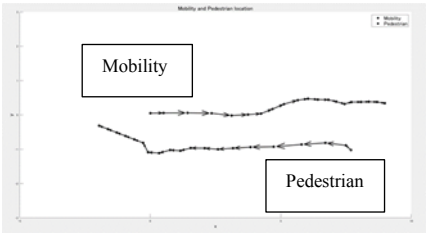


Figure 9: Trajectory example of without behavior prediction path planning.

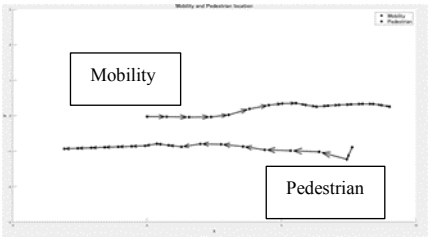


Figure 10: Trajectory example of small avoidance cooperative path planning.

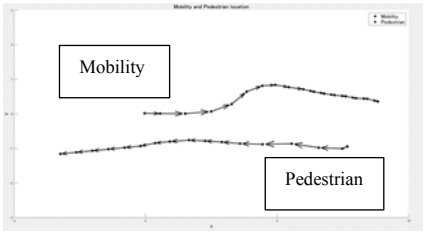


Figure 11: Trajectory example of large avoidance cooperative path planning.

The paths also showed that large avoidance was avoided to a greater extent than small avoidance and that the timing of avoidance was delayed without the collision risk area.

3.2 Subjective Evaluation Results

Subjective evaluation of pedestrian avoidance performance was carried out on nine subjects. Two trials of each condition were made to each subject. Experiments were carried out based on the approval of the Human Ethics Review Committee of Kanagawa institute of technology.

Subjective evaluation was performed with 4 evaluation items. They are "distance from the autonomous mobility when passing by", "speed of the mobility when passing by", "smoothness of passing by (avoidance performance)", and "reliability when passing by". These items were evaluated in five levels, with the following ratings: 'good', 'a little good', 'undecided', 'a little bad', and 'bad'.

Subjective evaluation results are as follows (Figs. 12 and 13). In the case of forward-facing pedestrians, the evaluation of all items improved in comparison without behavior prediction and small avoidance. In the case of downward-facing pedestrians, the evaluation values of all items improved in comparison without behavior prediction and large avoidance. Especially, in the case of downward-facing pedestrians, the evaluation of distance was considerably improved. No

change was observed in the evaluation of speed, partly because the vehicles were driven at the same speed in all three conditions.

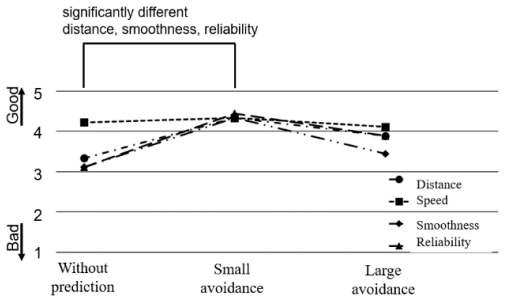


Figure 12: Subjective evaluation results for conscious pedestrians.

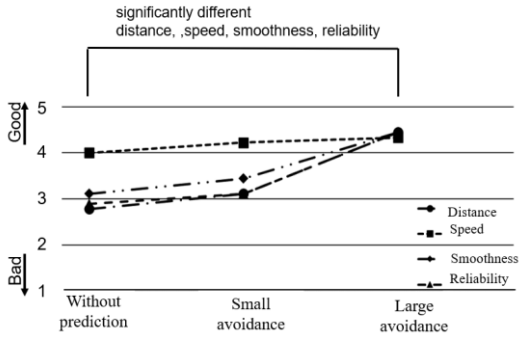


Figure 13: Subjective evaluation result.

Significant difference tests were conducted paired t-test with a significance level of 0.05 and a null hypothesis of 'no difference in mean values between the two groups. In this study, subjective ratings were treated as interval scales. Without behavior prediction and small avoidance were compared for forward-facing pedestrians, and without behavior prediction and large avoidance were compared for downward-facing pedestrians. The p-values for each item are shown below (Tables 3 and 4).

Table 3: Result of significance test for forward facing pedestrian (* $p<0.05$).

item	P-value
Distance	0.02* significantly different
Speed	0.173
Smoothness	0.046* significantly different
Reliability	0.021* significantly different

Table 4: Result of significance test for downward facing pedestrian (* $p<0.05$, ** $p<0.01$).

item	P-value
Distance	0.005** significantly different
Speed	0.040* significantly different
Smoothness	0.014* significantly different
Reliability	0.006** significantly different

4 DISCUSSION: THE RESULTS OF THE SUBJECTIVE EVALUATION

Both forward and downward-facing pedestrians improved the subjective evaluation results for all items except speed. Speed was highly rated in all three conditions, with no significant differences observed. It is considered that this is because the mobility moved at a constant speed under all experimental conditions.

In the subjective evaluation of forward-facing pedestrians, the evaluation of large avoidance was slightly worse than that of small avoidance. Several subjects commented on the poor smoothness of large avoidance, such as "I felt poor smoothness" and "If I were a human, I would feel un-comfortable as if I were being large avoided". From this result, it seems that small avoidance is appropriate for for-ward-facing pedestrians.

Significant differences in distance and reliability were found for forward-facing walking. This result is thought to be due to the fact that the mobility without prediction (no collision risk area) pass pedestrians at a close distance, while those with a collision risk area maintain a certain distance while avoiding pedestrians.

In downward-facing walking, significant differences were observed in all items except speed. In forward-facing walking, the presence of mobility can be confirmed early on in the effective field of view, whereas in downward-facing walking, the effective field of view is narrower than in forward-facing walking because walking is done while gazing at the smartphone [9], and the pedestrian only confirms the presence of the mobility when it enters the peripheral field of view just before passing by. Therefore, the evaluation of distance, the reliability and smoothness decreased, whereas with the collision risk area, the reliability also improved because the robot maintains a maximum safe distance and makes a larger avoidance compared to forward-facing walking.

5 CONCLUSION

In this study, we proposed a path planning algorithm that adapts to the face direction of pedestrians and safely avoids pedestrians who are walking while on their smartphones.

The effectiveness of this algorithm was confirmed by subjective evaluation.

This method would enable the operation of advanced collaboration between pedestrians and autonomous mobility on campus.

REFERENCES

[1] D. Fox, W. Burgard, and S. Thrun, “The dynamic window approach to collision avoidance,” IEEE Robot. Autom. Mag., vol. 4, no. 1, pp. 23-33,(1997).
[2] S. M. LaValle, “Rapidly-Exploring Random Trees: A New Tool for
Path Planning,” Tech. Rep., (1998).
[3] Takumi Goto, “Pedestrian behavior prediction and forecast circle generation using Kalman filter”(2019)
[4] Hiroshi Hisahara et al, “Human Avoidance Function for Robotic Vacuum Cleaner Through Use of Environmental Sensors - Roomba® Making Way for Humans”, IEEE, Fifth International Conference on Intelligent Systems, Modelling and Simulation, pp.64-67.(2014)
[5] Dongxiang Zhang, Zongjun Xie et al, “Real-Time Navigation in Dynamic Human Environments Using Optimal Reciprocal Collision Avoidance”, IEEE International Conference on Mechatronics and Automation.(2015).
[6] J.Redmon et al., “You Only Look Once: Unified, Real-Time Object Detection”, 2016 IEEE Conference on Computer Vision and Pattern Recognition (2016).
[7] “Plan Mobile Robot Paths Using RRT”, https://jp.mathworks.com/help/nav/ug/plan-mobile-robot-paths-using-rrt.html.
[8] Chisato Shibata et al, “Driver characteristic estimation using vehicle behavior fata while preceding vehicle decelerating”, Information Processing Society of Japan (2016).
[9] Daisuke Saito “Change in effective visual field using smartphone with walking”, Biomedical Fuzzy System Association (2019).

Title: Pedestrian Cooperative Autonomous Mobility -Path planning adapted to pedestrian face direction-

Author: Yuto Yada, Shunsuke Michita, Seiji Komiya , Toshihiro Wakita
Corresponding Author: Yuto Yada
Tel: 046-291-3091
Fax: 046-242-6806
E-mail: s2284001@cco.kanagawa-it.ac.jp

Implementation of Spatio-temporal Fencing for Crowd Sensing in a Smartphone Applications

SHOTA SUZAKI†, NOBUHITO MIYAGAWA†, and KATSUHIKO KAJI†

†Aichi Institute of Technology, Japan

Abstract - Crowdsensing platforms are available to solve the problems of dedicated system development and operating costs associated with crowdsensing. Among them, there are several crowdsensing platforms that provide incentive elements to motivate and maintain users. We build a crowdsensing platform (Lavlus) to mitigate disincentive factors that can be used in conjunction with incentive factors. This study implements a smartphone application for Lavlus. The Lavlus smartphone app needs to have the following functions: sensing project download, sensing request notification, automatic sensing, and sensor data upload. Among them, sensing request notification and automatic sensing are based on Spatio-temporal fencing. Geofencing used latitude and longitude, which are easy for users to visually recognize, but GPS is unstable indoors and in areas with many buildings. Therefore, a margin is provided on the geofence to determine whether the user has entered or exited the geofence with certainty. Lavlus may also target specific facilities, such as amusement parks or factories, as actual use cases. In this case, the geofence may be arbitrarily polygonal. We then draw a circle with the user at its center and generate geofencing points at eight points on the circumference of the circle. If one or more of the generated points are inside the geofence, the system determines that it is about to enter space-time and issues a notification whether or not it will cooperate with sensing. If all eight points are inside, the application judges that it has definitely entered the area and performs automatic sensing. If all eight points are outside, the application judges that it has definitely exited and terminates sensing. This allows proper geofencing even when the geofence is arbitrarily polygonal. In this paper, we implemented these functions and verified their operation. As a result, the geofence worked properly for arbitrary polygons.

Keywords: Spatio-temporal fencing, crowdsensing, crowdsensing platform, smartphones

1 Introduction

In recent years, the number of smartphones equipped with sophisticated sensors has been increasing, and a wide variety of sensors are becoming available. Therefore, there are a number of research that use smartphone sensors[1][2]. Crowdsensing is an attempt to utilize the sensing capability of smartphones[3][4][5]. Crowdsensing is currently being employed in research and surveys[6][7][8]. The implementation of crowdsensing requires the development of a dedicated system, which is expected to incur significant initial and running costs. In addition, in order to encourage many collaborators to cooperate in crowdsensing, there are issues such as reducing the time and effort required for collaborators and eliminating their anxiety. As for the time and effort required for collaborators, the burden of operating and communicating with smartphones can be mentioned. The concerns of the collaborators include worries about providing data due to privacy barriers of the collaborators and distrust of sensing. Furthermore, since crowdsensing handles a lot of sensitive data such as sensor data, security and privacy protection measures are essential.

The initial and running costs required to implement crowdsensing can be solved by a crowdsensing platform. However, crowdsensing platforms face several challenges, which we have addressed by proposing a crowdsensing platform based on Spatio-temporal fencing called ”Lavlus”. Create a crowdsensing platform to enable easy use of crowdsensing and diverse data collection to significantly reduce costs (time, money, and effort) in research and surveys. We also propose Spatio-temporal fencing, and expect those who use crowdsensing to collect sensor data (hereafter referred to as ”requesters”) to easily define the scope of sensing, and those who cooperate with crowdsensing to provide sensor data (hereafter referred to as ”collaborators”) to have a clear perception of the space-time that is being sensed. When using Lavlus, the requester uses a dedicated web application and the collaborator uses a dedicated mobile application. This study is about mobile applications of Lavlus.

2 Related Research

There are several research that use crowdsensing to collect data from a large number of people for estimation and analysis. For example, crowdsensing using mobile devices is used to collect and share ambient sounds to conduct noise surveys[6][7], and motion sensors such as accelerometers are used to collect data from car users to estimate road conditions such as icy and paved roads, and road geometry such as flat and hollow surfaces[8]. In these research, the development of a crowdsensing system is expected to incur significant costs. To implement crowdsensing, it is necessary to develop a sensing smartphone application exclusively for the collaborator and a server to manage the collected data.

Therefore, a crowdsensing platform would be very useful if the requester wishes to use crowdsensing to collect data. For the requester, it is no longer necessary to create and distribute a dedicated smartphone application for sensing for each research, thus eliminating the time and effort spent on these tasks. As for the collaborators, there is no need to install a separate smartphone application for each research, and there is no need to use separate applications for each research. In addition, the use of a common smartphone application can lead collaborators to other crowdsensing applications, which can lead to the acquisition of many collaborators.

Next, we discuss related research on crowdsensing platforms. There are already some that operate as simple platforms mainly for the purpose of reducing system development costs[9][10]. However, securing collaborators is very important for crowdsensing platforms. Related researches include those to improve and maintain the motivation of collaborators and to secure collaborators.

This includes research using monetary incentives[11] and research that uses gamification to provide non-monetary incentives[12]. There are also research that offer a flexible choice between monetary and non-monetary incentives. In this research, our approach is to remove disincentives for collaborators through Spatio-temporal fencing, and we are considering introducing a mechanism that can provide incentives in the future.

Next, we discuss related research on crowdsensing platforms. Since securing collaborators is extremely important for a crowdsensing platform, it is necessary to devise ways to increase and maintain motivation. There are two ways to motivate collaborators: research [11] using monetary incentives and research [12] using non-monetary incentives. These methods of increasing and maintaining motivation focus on incentive factors, and not much is said about disincentive factors. This system aims to reduce the disincentive factors that can be used in conjunction with these incentive factors.

Smartphones are used for many of the sensing terminals in the crowdsensing platform[9][10][11]. One of the advantages of using smartphones as sensing terminals for crowdsensing is that smartphones have high penetration rates and there is no need to secure, distribute, or upgrade new sensing terminals. In addition, if a mobile application dedicated to crowdsensing is created, people around the world can participate in crowdsensing via the Internet. However, as a disadvantage, smartphones are likely to be carried at all times, and the sensor data obtained may violate privacy if the collaborator senses at unintended places and times. In this study, Spatio-temporal fencing is expected to provide sensing that does not violate the privacy of collaborators.

3 Crowdsensing Platform Based on Spatio-temporal Fencing

This section describes the details of a crowd-sensing platform (Lavlus) based on Spatio-temporal fencing. Each section is organized in the following order: definition of space-time fencing, overall view of Lavlus

3.1 Definition of space-time fencing

Spatio-temporal fencing is defined as ”a fencing method that extends geo-fencing by adding a time element. An overview of Spatio-temporal fencing is shown in Figure 1. Geo-fencing is a fencing method that uses location-estimation technologies such as GPS, Wi-Fi, and BLE beacons to generate virtual boundaries, and provides specific services when a user enters or exits the boundary. With the increasing accuracy of location estimation and the widespread use of devices such as BLE, which can easily construct geofences, various applications using geofencing have been realized. In other words,

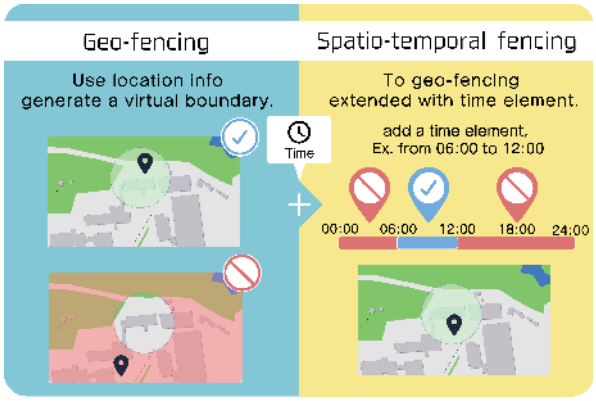


Figure 1: Spatio-temporal fencing with additional geo-fencing and time elements

Spatio-temporal fencing is a fencing method that generates virtual boundaries by specifying time and space, and performs a specific service when a user enters or exits the boundary. The specific service is ”sensing” in this study. Sensing is initiated when it enters within the virtual boundary and terminated when it exits outside the boundary.

One advantage of Spatio-temporal fencing is that delimiting the boundaries by time and area allows the requester to specify various situations for crowd-sensing. The range that can be defined for Spatio-temporal fencing is the range within which time and space can be properly separated. For example, a park from 3 p.m. to 5 p.m. or an amusement park during opening hours. On the other hand, the extent to which time and space cannot be properly separated cannot be defined. For example, it is not suitable only during rainfall, when time cannot be properly separated, or on a moving train, when space cannot be properly separated.

3.2 Overall view of Lavlus

The overall view of Lavlus is shown in Figure 2. The sequence of steps in Lavlus is ”defining a sensing project in a web app,” ”Spatio-temporal fencing,” ”accepting a sensing request,” ”automatically sensing,” ”automatically uploading under Wi-Fi,” and ”using the data. The requester defines the details of the sensing request in detail on the project management web application and creates a sensing project. The collaborator downloads the sensing project created by the requester. Spatio-temporal fencing is performed by the collaborator according to the sensing project, based on the definition in the 3.1 clause, and notifications are sent only if the collaborator is in the set space-time. If the collaborator agrees with the requestor and the contents of the crowdsensing presented on the sensing request screen, and cooperates with the sensing, the collaborator accepts the sensing request. When a collaborator accepts the sensing request and enters the time-space, sensing starts automatically in the background. When the sensing is finished and the collaborator is under Wi-Fi, the sensor data is uploaded.

This platform can be used only for crowd-sensing, where space-time can be set appropriately and sensing is done in unconsciousness. For example, suppose that an amusement

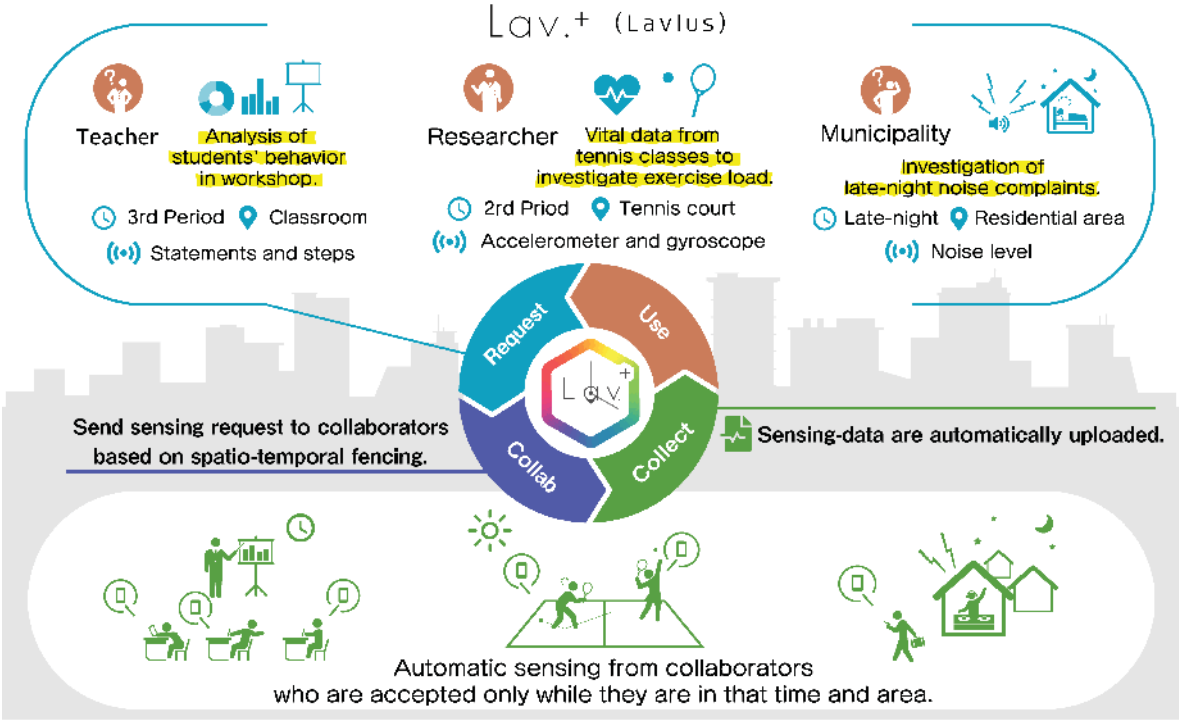


Figure 2: Overall view of Lavlus

park management company tries to sense the movement history of amusement park visitors in order to understand their trends. In this case, the time is set from the opening time to the closing time of the amusement park, the space is set to the inside of the amusement park, and the required sensor data are position information and acceleration. As another example, suppose a teacher wants to make environmental measurements to control the air conditioning of a classroom where he or she teaches a class. In this case, the time is set to the time of day when the class is held, the space is set to the classroom, and the required sensor data are temperature and humidity.

4 Application for automatic sensing upon entry into a specific time-space

This section describes an application for automatic sensing upon entry into a specific time-space. In section 4.1, we first define the requirements specification for Lavlus' mobile application. Section 4.2 describes the implementation of an application that automatically senses when entering a specific time-space. Section 4.3 describes the implementation of Spatio-temporal fencing. Section 4.4 describes the implementation of sensing request notifications. Section 4.5 describes the implementation of automatic sensing.

4.1 Requirements specification for Lavlus' mobile application

The Lavlus mobile application requirements specification is shown in Figure3. The Lavlus mobile app needs to be able to request download, sensing request notification, automatically sensing and sensor data upload.

The Lavlus mobile app needs to reduce the disincentive factor for collaborators. In order to reduce the physical costs for the collaborator, it is necessary to notify the collaborator and reduce the collaborator's own operations, and not to overwhelm the data traffic of the terminal. Notification should be issued only to collaborators who are likely to enter the space-time to minimize notification to collaborators. In addition, once a collaborator accepts or rejects a sensing request, no notification needs to be issued from that project. The requester may feel annoyed if notifications are issued for crowdsensing that he/she cannot participate in or for crowdsensing that he/she once rejected. Therefore, notification to collaborators should be kept to a minimum. To reduce the number of operations by the collaborator, the collaborator's operations should be limited to tapping the sensing request notification and tapping the Accept/Reject button on the sensing request screen, with the exception of installing applications, etc. Therefore, all Spatio-temporal fencing and sensing must be done automatically in the background. Collaborators may stop cooperating with crowdsensing if there is too much manipulation or effort involved before they cooperate. When downloading a sensing project, it is necessary to download only those sensing projects in which collaborators are likely to participate in order not to overwhelm the communication and data capacity of the terminal. It also needs to be uploaded automatically when connected to Wi-Fi after sensing is complete.

To reduce the psychological cost for collaborators, they need to be able to delete sensor data that they do not want to be used. It is also necessary to be able to delete sensor data that you do not want to be used even after uploading. Therefore, collaborators need to be able to reject a sensing project

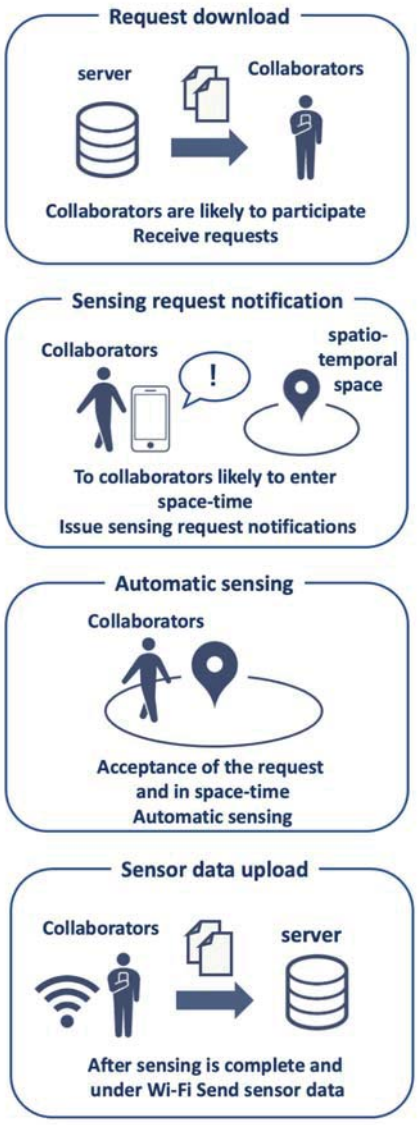


Figure 3: requirements specification for Lavlus' mobile application

that they have already accepted, or request the deletion of sensor data that they have already uploaded. If the data cannot be deleted once sensed, collaborators cannot casually sense the data. In this case, the number of collaborators may decrease because it will be difficult for them to participate in crowdsensing.

4.2 Implementation of an application that automatically senses when entering a specific time-space

Implement a mobile app for sensing corresponding to the sensing project produced by the requester. This application only implements Spatio-temporal fencing, sensing requests, and automatic sensing out of those described in section 4.1. The overall diagram of the implemented application is shown in Fig4.

Issue a sensing request notification to collaborators who are likely to enter the space-time or who are already inside the space-time. Tap the sensing request notification to display the sensing request screen. The collaborator checks the information about the requester and crowdsensing presented on the sensing request screen and chooses whether or not to cooperate with crowdsensing. If the collaborator agrees to cooperate with crowdsensing, he/she presses the "Accept" button on the sensing request screen; if not, he/she presses the "Deny" button. If the collaborator agrees to crowdsensing and is definitely inside the space-time, it is automatically sensed.

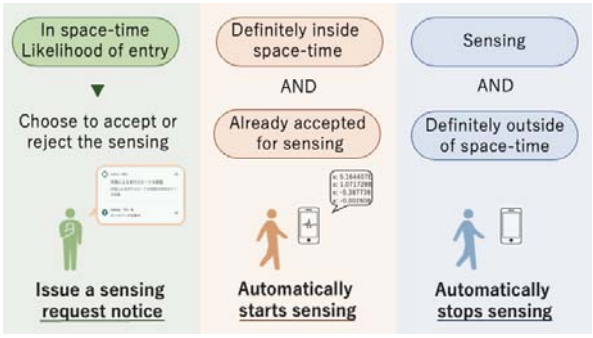


Figure 4: Overall view of an application for automatic sensing upon entry into a specific time and space

4.3 Implementation of Spatio-temporal fencing

Latitude and longitude, BLE beacons, and Wi-Fi are some of the methods used to generate geofences. Lavlus needs to ensure that the requester and collaborator are aware of the geofence. For this reason, the latitude and longitude were used in this case because they are visually easy to recognize.

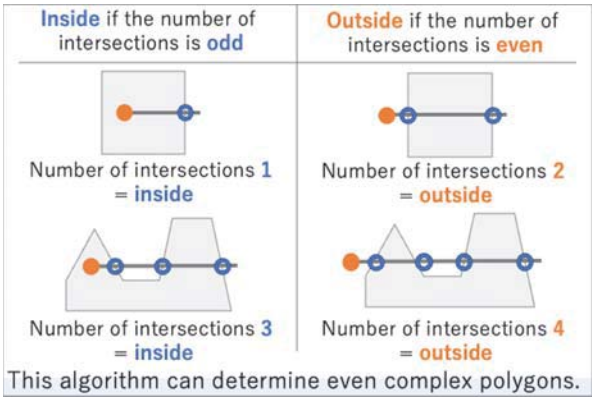


Figure 5: Geo-fencing using point inside/outside determination

Use the polygon interior/exterior decision algorithm to deal with the case where the geofence is an arbitrary polygon. The inside/outside judgment for a polygon of points is shown in Figure5. To determine whether a point is inside or outside a polygon, first draw a straight line from the point (red circle) to the polygon. If the number of intersections (blue circles)

between the line and the polygon is odd, the line is inside the polygon; if it is even, the line is outside the polygon. In this case, if exactly a line overlaps an edge of a polygon or a line overlaps a vertex of a polygon, a wrong decision is made. However, since latitude and longitude have seven decimal places and location information keeps changing, exactly overlapping cases are not considered this time. No matter how complex a polygonal area the requester sets up, such as the lower part of Figure5, it is possible to determine whether it is interior or not. Although not implemented in this case, it is also possible to determine the area in the shape of a circle. In addition, since only GPS is used to determine the area, only a flat area can be determined in this case. Determination of three-dimensional areas such as the first and second floors is a subject for future work.

In order to cope with various situations, such as sensing only when the user is definitely inside the space-time interval, or issuing sensing request notifications to collaborators who are likely to enter the space-time interval, we have implemented a margin that allows the space-time interval to be expanded and contracted. If the space-time is to be determined with certainty, the space-time is reduced by a margin. When determining whether or not the user is in the space-time perimeter, an expanded space-time margin is used. If the geofence is a rectangle made up of only lines parallel to the latitude and longitude lines, the margin can be implemented using addition and subtraction. To enlarge the geofence, add a margin to the northernmost latitude and easternmost longitude, and subtract a margin to the southernmost latitude and westernmost longitude. To reduce the geofence, subtract a margin from the northernmost latitude and easternmost longitude and add a margin to the southernmost latitude and westernmost longitude. However, this application needs to deal with the case where the geofence is an arbitrary polygon. So we let the collaborator, not the geofence, have the margin.

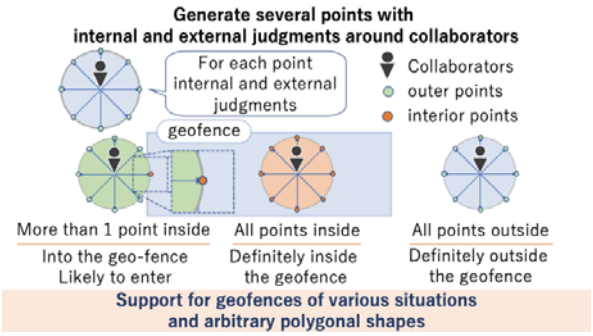


Figure 6: Based on a few points on the circumference centered on the collaborator, the algorithm determines if the collaborator is likely to enter the geofence, if the collaborator is definitely inside the geofence, or if the collaborator is definitely outside the geofence.

A summary of the margins that the collaborator has is shown in the figure. First, draw a circle with a radius of the length of the margin centered on the collaborator. It then generates several points equally on its circumference. Geofencing is performed on all the generated points. If one or more of the generated points are inside the geofence, the collaborator is considered likely to enter the geofence. If all generated points are inside the geofence, the collaborator is definitely inside the geofence. If all generated points are outside the geofence, the collaborator is definitely outside the geofence.

4.4 Implementation of sensing request notifications

To reduce the number of sensing request notifications to collaborators, sensing request notifications are issued to collaborators who are most likely to participate in crowdsensing. Collaborators who are close to the space-time set for the sensing project and already inside the space-time are considered likely to participate in that crowdsensing project. For example, a collaborator who is in the vicinity of crowdsensing that has already begun, or a collaborator who will soon begin crowdsensing where he or she is located. Therefore, we take a margin to widen the time-space, and issue a notification if the user is inside that time-space. Assume that collaborators who are located far from the space set aside for the sensing project are unlikely to participate in that crowdsensing. For example, collaborators who are far away from the crowdsensing that has already begun. Issue a notice to a collaborator who is unlikely to participate in crowdsensing, and even if they accept the sensing request, they are unlikely to receive sensor data. In addition, collaborators may feel uncomfortable with repeated issuance of crowd-sensing notifications in which they do not participate.

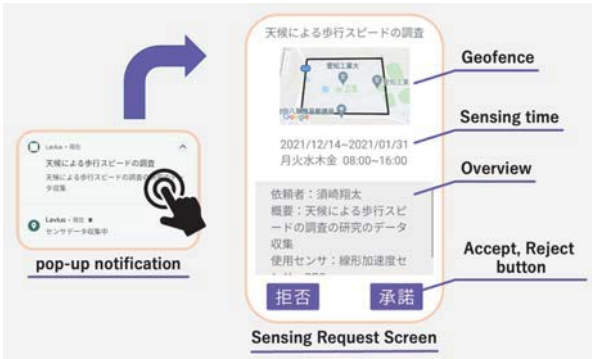


Figure 7: Example of issued notification and sensing request screen

When the collaborator taps the sensing request notification, the sensing request screen is launched and the requester's name, the sensor to be used, and the time and space are presented (Figure7). If the collaborator is satisfied with the information presented and decides to cooperate, the collaborator presses the "Accept Sensing" button and cooperates with crowdsensing. The collaborator confirms the information presented on the sensing request screen. Examples of information presented on the sensing request screen include the requester's name, the name of the sensor to be used and its frequency. At this time, collaborators can reject crowdsensing by pressing the "Reject Sensing" button if they have any distrust or disagreement with the information presented.

4.5 Implementation of automatic sensing

To reduce the manipulation of the collaborator, sensing is automatically performed in the background when the collaborator has definitely agreed to the sensing request inside the space-time. Therefore, the collaborator does not need to consciously open the smartphone application, and the sensing does not terminate during the process unless the application itself is terminated. In addition to reducing the burden of the collaborator's operations, we expect the system to sense the user's usual behavior because it does not make the user aware of the sensing process.

In order to determine whether or not the collaborator is definitely inside the space-time of the collaborator, a narrower space-time margin is taken on entry and a wider space-time margin is taken on exit. If the geofencing is near the boundary and the location information becomes unstable, the entry/exit decision is repeated. If the location information is unstable around space-time, it may be determined to be inside space-time even though it is outside space-time. To prevent this, a margin is taken. However, if the space-time is small, too narrow a margin makes it difficult or impossible to enter the space-time. For example, for a geofence that is less than 10 meters in either direction, a narrowing margin of 5 meters or more will make it impossible to enter. Similarly, if the margin to narrow the time is more than the set time, entry will not be possible. Therefore, when narrowing the space-time interval, the margin must take into account the size of the space-time interval. At this time, margins are set for each sensing project.

As a crowdsensing platform, it supports many sensors and free frequencies, and abstracts sensor data that would violate privacy. Many sensors and frequencies need to be supported to accommodate a variety of crowd sensing. For example, gait estimation may require acceleration sensors, barometric pressure sensors, angular velocity sensors, and location information. Sound sensors may be needed for noise measurements, and temperature and humidity for environmental measurements. Not only the type of sensor but also the frequency that can be freely set is necessary. The frequency of the same sensor varies depending on the application. For example, when a barometric sensor is used for gait estimation, it senses at a high frequency. On the other hand, weather estimation does not require sensing at high frequencies, so sensing is done at lower frequencies. The type of sensor and the frequency for each sensor are set in the sensing project created by the requester. This application allows sensing for multiple sensors and sensor frequencies according to the sensing project. For example, frequencies can be set on a per-sensor basis, such as sensing acceleration at 50 Hz and barometric pressure at 10 Hz. It may also enter multiple time-spaces simultaneously and participate in multiple crowdsensing activities. This application manages sensor types and their frequencies for each sensing project, so that multiple sensing can be performed simultaneously. For example, even when cooperating simultaneously with crowdsensing that takes acceleration at 50 Hz and crowdsensing that takes acceleration at 20 Hz, it is possible to sense appropriately.

5 Experiment

We verified that the geofence works properly when it is arbitrarily polygonal, including the margins. The evaluation items for the Experiment are shown in Fig.8. First, check to see if notifications have been issued to collaborators who are likely to enter the geofence. The collaborators most likely to enter the geofence are as described in the4.4 section. As described in section4.3, when one or more points with an inside/outside decision attached to a collaborator are inside a geofence, the geofence is considered to have entered an expanded geofence and a notification is issued. Next, verify that the sensing starts when you are definitely inside the geofencing. If it is working properly, when all the points with inside/outside judgments attached to the collaborator enter the geofence, it judges that it is definitely inside the geofence and starts sensing. Finally, check to see if the sensing is terminated when the geo-fence is definitely exiting the geofence. If it is working properly, when all the points with inside/outside judgments attached to collaborators have left the geofence, it is determined that they are definitely outside the geofence, and sensing is terminated. The experimental environment for Experiment and the actual route moved are shown in Figure9. A polygonal geo-fence is set up as shown in Figure8, and the geo-fence is moved through (Figure8 orange arrows).

Figure 8: Experiment of geofencing operation

The results of the operation verification are shown in Fig10. First, the issuance of the notice was verified upon entry into the expanded geofence. Next, we verified that sensing begins upon entry into the reduced geofence. Finally, we confirmed that sensing was terminated when they exited the expanded geofence. From the above, it can be said that proper operation was confirmed.

6 Conclusion

In this paper, we defined the requirements for the functions required for mobile applications in a crowdsensing platform. From these, Spatio-temporal fencing, sensing requests, and automatic sensing were implemented. The implementation of Spatio-temporal fencing used polygon inside/outside decisions to accommodate the case where the geofence is an arbitrary polygon. In order to cope with various situations, such as sensing only when the user is sure to enter space-time, or issuing sensing request notifications to collaborators who are

Figure 9: Verification of geofencing operation

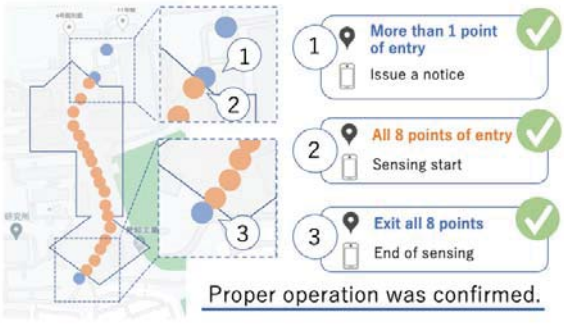


Figure 10: Results of geofencing operation verification

likely to enter space-time, the system implements a margin that allows expansion and contraction of space-time. In addition, to accommodate the case where the geofence is a complex rectangle, a margin is provided for the location information of collaborators. The first step in implementing a sensing request was to define collaborators who are likely to participate in the sensing request. Notification is then issued only to collaborators who are likely to participate in the sensing request. In addition, a sensing request screen was implemented to determine whether or not the collaborator would participate in the sensing. The company agreed to the sensing request and implemented a function that automatically senses when the user enters the Spatio-temporal space. In order to reliably determine whether a collaborator enters or exits space-time, a narrow space-time margin is used for entry, and a wide space-time margin is used for exit. As a crowdsensing platform, it supports many sensors and free frequencies, and abstracts sensor data that would violate privacy.

In the operation verification, we checked whether the Spatio-temporal fencing was properly performed. As a result, Spatio-temporal fencing was performed properly. In addition, we verified whether the sensing was appropriate for actual use cases. As a result, various sensor data were collected.

Future work includes the implementation of necessary functions for mobile applications in a crowdsensing platform based on Spatio-temporal fencing, which was not implemented in this study. It is necessary to implement the linkage with the server such as sensing project download and sensor data upload, which could not be implemented this time.

One of the challenges of Spatio-temporal fencing is that it uses GPS for geo-fencing, which makes the operation unstable indoors where GPS accuracy is degraded. Since only latitude and longitude are used for geofencing at this time, geofencing cannot be performed properly when GPS blurring is more severe than expected, such as indoors or in areas with many high-rise buildings. Latitude and longitude alone cannot be used to determine the first or second floor. Therefore, in addition to geofencing in latitude and longitude, it is necessary to implement BLE and Wi-Fi.

REFERENCES

[1] Suyama, A., Inoue, U.: Using geofencing for a disaster information system, 2016 IEEE/ACIS 15th International Conference on Computer and Information Science, pp.1-5, 2016.

[2] Sugimori, D., Iwamoto, T., Matsumoto, M.: A Study about Identification of Pedestrian by Using 3-Axis Accelerometer, 2011 IEEE 17th International Conference on Embedded and Real-Time Computing Systems and Applications, Vol.2, pp.134-137, 2011.

[3] Burke, J., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., Srivastava, M. B.: Participatory Sensing, Workshop on World-Sensor-Web: Mobile Device Centric Sensor Networks and Applications, 2006.

[4] Raghu, G., Fan, Y., Hui, L.: Mobile Crowd Sensing: Current State and Future Challenges, IEEE Communications Magazine, Vol.49, pp.32-39, 2011.

[5] Liu, J., Shen, H., Zhang, X.: A Survey of Mobile Crowdsensing Techniques: A Critical Component for the Internet of Things, 2016 25th International Conference on Computer Communication and Networks, pp.1-6, 2016.

[6] Eiman, K.: NoiseSPY: A Real-Time Mobile Phone Platform for Urban Noise Monitoring and Mapping, MONET, Vol.15, pp.562-574, 2010.

[7] Nicolas, M., Matthias, S., Bartek, O.: Participatory noise pollution monitoring using mobile phones, Information Polity, Vol.15, pp.51-71, 2010.

[8] Bin, P., Kenro, A.: Detecting the road surface condition by using mobile crowdsensing with drive recorder, 2017 IEEE 20th International Conference on Intelligent Transportation Systems, pp.1-8, 2017.

[9] Tangmunarunkit, H., Hsieh, C. K., Longstaff, B., Nolen, S., Jenkins, J., Ketcham, C., Selsky, J., Alquaddoomi, F., George, D., Kang, J., Khalapyan, Z., Ooms, J., Ramanathan, N., Estrin, D.: Ohmage: A General and Extensible End-to-End Participatory Sensing Platform, Association for Computing Machinery, Vol.6, No.3, 2015.

[10] Ferreira, D., Kostakos, V., Dey, A. K.: AWARE: Mobile Context Instrumentation Framework, Frontiers in ICT, Vol.2, pp.6, 2015.

[11] Jayarajah, K., Balan, R. K., Radhakrishnan, M., Misra,

A., Lee, Y.: LiveLabs: Building In-Situ Mobile Sensing & Behavioural Experimentation TestBeds, MobiSys 16: Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services, pp.1-15, 2016.

[12] Shogo, K., Yuki, M., Hirohiko, S., Manato, F., Yutaka, A., Keiichi, Y.: Gamified Participatory Sensing in Tourism: An Experimental Study of the Effects on Tourist Behavior and Satisfaction, Smart Cities, Vol.3, No.3, pp.736-757, 2020.

Sleep Effects on Noise.

–Comparison of Sleep Effects between Students and Working People-

Tatsuki Takahashi*,Masaaki Hiroe** , and Mari Ueda*

*Kanagawa Institute of Technology, Japan
**Kobayasi Institute of Physical Research, Japan
s2285013@cco.kanagawa-it.ac.jp

Abstract - We have investigated how noise affect sleep using recorded transportation noise in participants' own home since 2019. In this study, we analyzed subjective and objective sleep assessment for twenty participants consist of each ten persons of working people and university students. The comparison between the two groups showed that the working peoples were more affected by noise-induced sleep in the subjective sleep assessment. In addition, it confirmed dose-response relationships between noise level and incidence of wake after sleep onset (WASO) derived from the objective sleep assessment differ from each traffic noise.

Keywords: sleep effects, traffic noise, questionnaire, body-movement, statistical analysis.

1 INTRODUCTION

Ensuring good quality sleep is a very important issue in today's world, where one out of every five Japanese citizens is becoming increasingly dissatisfied with their sleep. Traffic noise is one of the environmental factors that affect sleep. The 2018 European WHO Environmental Noise Guidelines [1] recommended that L_{night} 40-45 dB (energy-averaged nighttime noise level) should be used to protect people's health, which became a major topic of discussion.

The authors also started an experimental study of noise-induced sleep effects in 2019, and in the previous year's study [2,3], the authors analyzed the association between subjective and objective sleep ratings and noise exposure regarding sleep effects. The results showed that subjective and objective sleep ratings did not agree and showed different trends. As a consideration of the cause of this difference, the previous study had conducted experiments on working peoples, but in the previous year's study, most of the participants were university students, and the difference in sleep habits was considered to be one of the reasons for the difference.

Therefore, in this study, experiments were conducted on 10 working peoples and 10 university students. Two evaluation methods were used: an objective sleep evaluation using an actigraphy body movement meter and a subjective sleep evaluation using two questionnaires.

1.1 Literature Review

Methods for studying the effects of noise on sleep include social surveys and laboratory experiments. The former method provides information on the effects of noise on sleep in real life, but the extent of the effects is based on subjective impressions, and it is difficult to grasp and control the noise environment during sleep. On the other hand, in the latter indoor experiment, it is possible to grasp and control the noise environment presented to the subject and observe the sleep effect objectively, but it is also necessary to consider the effect of changes in the sleep environment on the subject. Therefore, in order to understand the effects on sleep corresponding to real life, in the experiment by Kaku et al., [4] traffic noise recorded at actual noise sites was presented to sleeping subjects using a CD boom box, and the effects on sleep were investigated using two methods: a hand-arm type body motion meter called an actigraphy and a questionnaire. The results confirmed that under conditions of equal L_{Aeq} presentation levels, railroad noise had a greater impact on sleep than road traffic noise. In addition, the actigraphy evaluation confirmed that the incidence of mid-sleep awakenings increased significantly for railroad noise when $L_{Aeq,1min}$ exceeded 50 dB, although there was not necessarily a clear relationship with the questionnaire.

2 METHODS

In this study, subjects were randomly presented with nine different test sounds, obtained from three types of traffic noise and three conditions of presentation level, and the effects of noise on sleep were determined by two methods: a questionnaire and an actigraph (wristwatch-type body movement meter: GT3X series, AMI). The actigraph used is shown in Fig. 1.

The questionnaire was administered as early as possible after waking up the next morning, asking whether and to what extent the test sound affected sleep. No sound was played on the first day of the experiment in order to get used to wearing the actigraphy, and the CD was played from the second day to the tenth night of the experiment.

The actigraph is an instrument that continuously measures the amount of body movement that occurs within a certain period of time and determines wakefulness based on a set sleep score algorithm. The Cole-Kripke method was used to determine wakefulness or sleep. These experimental methods have been approved by the ethical review committee of the University.

Among the participants in the experiment, three working peoples and two students had their noise exposure increased by +5 dB, although the position of the CD player and the type of sound source were the same. A sound level meter (RION, NL-52) was placed at the bedside during the

experiment, and room background noise and presentation levels were measured continuously for 10 days. The sound level meter used is shown in Fig. 4.

Fig.1 Actigraph (GT3X series, AMI)

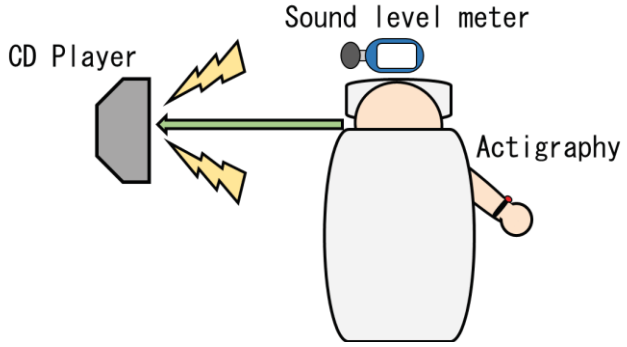


Fig 2 Diagram of the experiment

Fig 3 CD player (Wave System series, BOSE)

Fig 4 Sound level meter (NL-52, RION)

2.1 STIMULI

The test sounds used as stimuli were road traffic noise[RTN], conventional traffic noise[CRN], and aircraft noise[ACN]. The road traffic noise was a fluctuating noise that always sounded and was always present during sleep. Railroad noise includes the sound of six trains running per hour. Aircraft noise includes four takeoff sounds of aircrafts per hour. Table 1 shows a list of the types of test sounds.

Table 1 Stimuli type and sound pressure level (SPL).

CD No.	Type of traffic noise	SPL ($L_{Aeq,1h}$)	SPL, +5 dB ($L_{Aeq,1h}$)
1	Road traffic noise (RTN)	42.5 dB	47.5 dB
2		35.0 dB	40.0 dB
3		27.5 dB	32.5 dB
4	Conventional railway noise (CRN)	42.5 dB	47.5 dB
5		35.0 dB	40.0 dB
6		27.5 dB	32.5 dB
7	Aircraft noise (ACN)	42.5 dB	47.5 dB
8		35.0 dB	40.0 dB
9		27.5 dB	32.5 dB

2.2 PARTICIPANTS

There were 10 working peoples (1 female) with a mean age of 36.8 years (standard deviation: 15.2 years) and 10 students (2 females) with a mean age of 21.4 years (standard deviation: 1.1 years).

2.3 SUBJECTIVE SLEEP ASSESSMENT

In this study, two types of web-based questionnaires were used in the subjective sleep assessment. One questionnaire was used to measure usual sleep sensation and sound sleep sensation. The usual sleep sensation questionnaire is shown in Table 2. The other questionnaire was administered as early as possible after waking up the next morning so that subjective sleep sensations would not be forgotten. The questionnaire for subjective sleep sensation is shown in Table 3.

Table 2 The questionnaire on usual sleep habit.

Q1	when do you usually enter your bed (bed or futon) and try to sleep (turn off the light)?
Q2	How much do the hours of lights-out vary from day to day?
Q3	What time do you usually wake up on average?
Q4	How much does your waking time change from day to day?
Q5	How much sleep do you usually get?
Q6	How much sleep do you get from day to day?
Q7	Do you get enough sleep?
Q8	What is the ideal amount of sleep for you?
Q9	Approximately how long do you usually stay in bed (bed or futon) before falling asleep?
Q10	Are you a good sleeper?
Q11	How many times per night do you wake up during the night (while sleeping)? (Please consider on

	average)
Q12	What triggers you to wake up most often? (Waking up in the middle of the night)
Q13	What triggers you to wake up most often? (Waking up in the morning)
Q14	How do you usually feel when you wake up in the morning?
Q15	How deep do you usually sleep?
Q16	Do you drink alcohol?
Q17	How is the environment around your house at night?
Q18	Do you hear any of the following sounds at your home at night?
Q19	Do you have trouble sleeping due to sound during your normal sleeping hours? (Please describe other sounds that cause you to have trouble sleeping)

Table 3 The questionnaire for subjective sleep sensation

Q1	What time did you get up this morning?
Q2	What time did you go to bed last night?
Q3	Were you able to sleep while listening to the CD?
Q4	What time did you stop the CD?
Q5	Did you wake up during sleep?
Q6-1	How many times did you wake up last night?
Q6-2	What time did you awaken from sleep?
Q7	How were the sounds heard during the experiment?
Q8	If you had trouble sleeping, what specifically was it that disturbed you?
Q9	Did you drink alcohol last night?
Q10	Did you take sleeping pills or cold medicine last night?
Q11	Did you sleep better than usual last night?
Q12	Did you sleep better last night than usual?
Q13	Overall, was your sleep good last night?
Q14	Do you feel more tired this morning than usual?
Q15	How was the depth of your sleep last night compared to usual?
Q16	How many dreams did you have last night?

2.4 OBJECTIVE SLEEP ASSESSMENT

The actigraph is a measuring instrument that measures the amount of body movement that occurs within a certain period of time in a series, and judges mid-arousal based on the occurrence of events that exceed a set threshold value.

Fig. 5 shows the method proposed by Kageyama et al. [5], for judging mid-onset arousal. As shown in the figure, one or two large body movements are not considered to be mid-arousal, but only three or more body movements under some rules are considered to be mid-arousal.

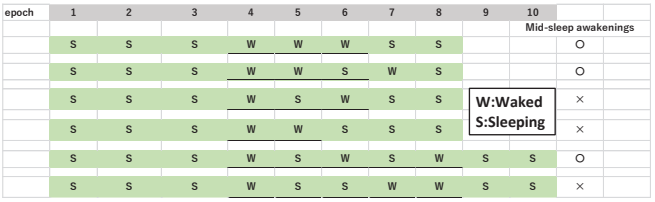


Fig.5 Diagram for determining mid-sleep awakenings

3 RESULTS

We compared the responses of working peoples and university students regarding their "usual sleep habits" (variation in waking and sleeping times, depth of sleep, etc.), which they were asked to answer of questionnaire shown in Table 2 (Mann-Whitney's U-test). The results showed that there are significant differences in Q1 (*p<0.017), Q2 (*p<0.019), Q3 (**p<0.007), Q4 (*p<0.041) and Q14 (*p<0.013). Thus, it can be said that working peoples went to bed and woke up earlier and lived more regularly (with less time fluctuation), but university students slept more soundly.

Using the answers to a self-administered questionnaire shown in Table 3 (subjective data) and wake/sleep judgment based on the amount of body movement (objective data), night-related outcomes were compared between the two groups (t-test). The results showed that there are significant differences in Q3 (**p<0.006), Q5 (***p<0.001), Q12 (**p<0.008), Q15 (*p<0.030) and Q16 (***p<0.001). The results showed that the working people group was the worse sleepers in Q.17. On the other hand, comparisons of objective data between the two groups showed there are significant differences in several outcomes as follows: Total Counts (*p<0.015), Total Minutes in Bed (**p<0.004), Total Sleep Time (*p<0.014) and Mean Wake Time (**p<0.004). Since the total counts was affected by sleep duration, it is possible that working peoples had a greater amount of movement than students. In addition, a significant difference was found in the mean awakening time for the subjective sleep evaluation item (Q5). However, although there was a significant difference in the mean awakening time between the two groups (2.8 min for working peoples and 2.2 min for students), considering the number of awakening events (15.4 times for working peoples and 16.4 times for students), the time spent in the awake state (= number of awakening events x mean awakening time) for both working peoples and students was 11-12% of total sleep time (362 min and 331 min). The mean number of awakenings was 43.1 min and 36.1 min, corresponding to 11-12% of the total sleep time (362 min and 331 min) for both working peoples and students.

The association between subjective and objective sleep ratings per night and presentation level $L_{Aeq,1h}$ was then tested separately for working peoples and students (Spearman's rank correlation test). The results are shown in Table 2, which indicates that the RTN participants were aware that noise exposure had an effect on their sleep state and arousal. None of the objective evaluation indices showed significant correlations with the presentation level.

Finally, Fig. 6 shows the relationship between $L_{Aeq,1min}$, and Wake After Sleep Onset (WASO) for the working

peoples and students. The correlation coefficients calculated assuming a simple increase in WASO with respect to onset level were significant for RTN (r=0.888*) and ACN (r=-0.879*) for working peoples and ACN (r=-0.934**) for students. clearly different from RTN (multiple comparisons of the difference in correlation coefficients using the Bonferroni method), but one should be cautious about negative correlations.

Table 4 Correlation coefficient between each subjective sleep assessment and sounds level $L_{Aeq,1h}$

	RTN	CRN	ACN
Q3	0.503** 0.033	0.351 0.305	0.264 0.247
Q5	0.218 0.039	0.295 0.413*	0.389* 0.433*
Q12	0.294 -0.065	0.237 0.130	0.257 0.228
Q15	0.124 -0.159	0.243 0.151	0.349 0.296
Q16	0.152 -0.206	0.116 -0.042	0.167 -0.120

Note: *p<0.05, **p<0.01, where p-value indicates the significance probability of correlation analysis. The upper value in each column is correlation coefficient for working people, and the lower one is that for students.

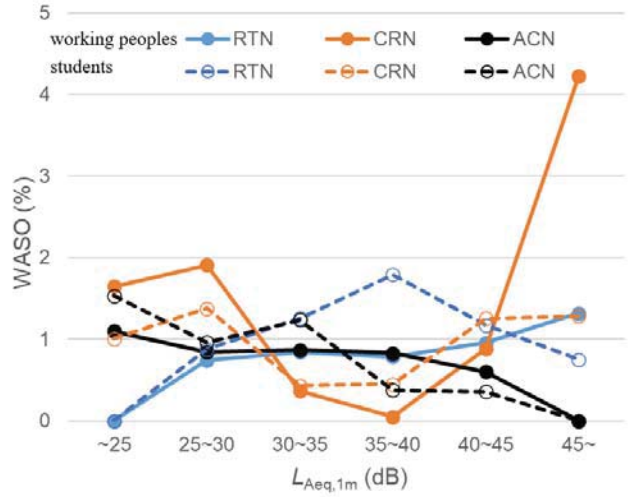


Fig.6 Dose-response relationship between SPL, $L_{Aeq,1min}$ and incidence of mid-sleep awakenings

4 DISCUSSION

From the results of this experiment, 10 working peoples and 10 university students with different sleep habits were exposed to traffic noise while sleeping, and their sleep states were evaluated by both subjective and objective sleep assessments. Subjective sleep evaluations revealed that working peoples were more likely to have regular sleep habits, with less fluctuation in their sleeping and waking times. However, the working peoples were found to have poorer sleep conditions. The objective sleep evaluation confirmed a significant difference in the average wake time of working peoples, but considering the number of awakenings, it was determined that there was no significant difference in the average wake time between university students and working peoples, which was equivalent to 11-12% of total sleep time (43.1 minutes for working peoples and 36.1 minutes for university students).

In addition, a comparison of subjective and objective sleep evaluations revealed a significant correlation between railroad noise and aircraft noise for working peoples in the questionnaire item about whether they were able to fall asleep while listening to CDs. In addition, significant correlations were found between the two groups of working peoples and university students in the questionnaire item of whether they were awakened during sleep and the presentation level. Thus, it was clear that the subjects were aware of being awakened during noise exposure. No significant correlation was found between the objective sleep evaluation and the presentation level.

Road traffic noise, railroad noise, and aircraft noise had different correlations with each other. Road traffic noise was positively correlated, railroad noise was uncorrelated, and aircraft noise was negatively correlated.

Comparison of subjective and objective sleep evaluations between university students and working peoples revealed that middle-aged and young working peoples had different results in terms of sleep onset and depth of sleep. These results were judged to be reasonable because they were considered to be due to the decline in physiological functions caused by aging, rather than to the effects of the working-age population.

REFERENCES

[1] World Health Organization Regional Office for Europe, Environmental Noise Guidelines for the European Region (2018).

[2] K. Nagai, F. Koyama, M. Ueda, H. Kasuga, and M. Hiroe, Sleep effects on noise – Follow-up test on sleep disturbance caused by traffic noise, 2020 Spring Meeting Acoustical Society of Japan, pp.593-594 (in Japanese).

[3] K. Enomoto, K. Nagai, F. Koyama, M. Ueda, H. Kasuga, and M. Hiroe, Sleep effects on noise part 2 – Discussions on the results of sleep analysis by actigraphy, 2020 Autumn Meeting Acoustical Society of Japan, pp.373-374 (in Japanese).

[4] J. Kaku, M. Hiroe, S. Kuwano, and S. Namba, Sleep disturbance by traffic noise, Jornal of sound and vibration, Vol.277, pp.459-464 (2004).

[5] T. Kageyama, T. Kurokawa and M. Kabuto, An epidemiologic study on insomnia in metropolitan areas IV: a relationship between incidence of waking estimated by actimetry and nighttime road traffic noise, Proceedings of the autumn meeting, the Institute of Noise Control Engineering of Japan, pp.65-68, 1997 (in Japanese).

Session 8:
IoT and Network
(Chair: Masashi Saito)

A Study on Application of Zero Trust Architecture to IoT Actuators

Nobuhiro KOBAYASHI*

*Department of Information Security, University of Nagasaki, Japan
nkoba@sun.ac.jp

Abstract -In recent years, the use of IoT systems in society has been progressing. As an example of the Society 5.0 that Japan is aiming for, it has been shown that information from sensors in physical space will be accumulated (big data) through IoT, analyzed by artificial intelligence (AI), and feedback to physical space to provide high-value-added information, suggestions, and control of devices. On the other hand, if cyber attacks on IoT systems cause service outages, unauthorized operations, or other problems, it could have a serious impact on our daily lives and economic and social activities. Therefore, in this paper, we especially consider security in IoT actuators that exert physical action as feedback from cyberspace to physical space. As a result of the study, we also present the Zero Trust IoT Security Framework, which applies the Zero Trust architecture to IoT actuators.

Keywords: IoT System, IoT Actuator, Cyber Security, Zero Trust Architecture, CPS

1 INTRODUCTION

In recent years, the use of IoT systems in society has been progressing. Society 5.0, which Japan is aiming for, is defined as "a human-centered society that achieves both economic development and solutions to social issues through a system that highly integrates cyberspace and physical space (the real world). As an example, information is collected from sensors in the physical space through IoT (big data), analyzed by artificial intelligence (AI), and fed back to the physical space with high-value-added information, proposals, and control of devices. On the other hand, if cyber attacks on IoT systems cause service outages, unauthorized operations, or other problems, our daily lives and economic and social activities may be seriously affected [1][2].

In this paper, we discuss, in particular, the results of our consideration of security in IoT actuators that exert physical action as feedback from cyberspace to physical space [3] [4][5][6][7]. Future directions for security measures in IoT actuators will also be presented.

2 ABOUT IOT ACTUATORS

The total number of IoT devices worldwide is expected to reach approximately 35 billion by 2022. In terms of sectors, "medical," "industrial," "consumer," and "automotive and aerospace" are expected to show high growth, replacing "telecommunications," which has been increasing [8].

The following features of this IoT device have a different impact on cybersecurity and privacy risks than traditional IT devices [4].

A) Transducer function

It interacts with the physical world and acts as an edge between the digital and physical environments. All IoT devices have at least one of two types of transducer functions.

Sensing function

Ability to measure and provide data on the physical world
Examples: temperature, optical sensing, audio sensing, radar

Actuator functions

Ability to create change in the physical world
Examples: heating coils, speakers, electronic door locks, drone operations, servo motors, robot arms

B) Interface Functions

Enable device-to-device interactions (device-to-device, device-to-human, etc.).

Application interface

Application programming interface, etc.

Human user interface

Functions for direct communication between IoT devices and humans
Examples: Touch screen, microphone, camera, speaker

Network interface

Ability of IoT devices to use communication networks
Example: Ethernet, Wi-Fi, LTE, ZigBee, etc.

C) Support Functions

The ability to support other IoT functions.
Examples: device management, cyber security features, privacy features

In this paper, A) IoT devices with actuator functions as transducer functions will be referred to as **"IoT actuators"**. IoT actuators are capable of influencing the physical world, which may lead to cyber attacks that may threaten human safety and life, damage and destroy equipment and facilities, and cause major disruptions such as the shutdown of critical services like social infrastructure. In Smart Cities for Society 5.0, a large number of IoT actuators will be deployed and utilized, and it is important to ensure their security.

3 SCOPE OF THE STUDY

Since IoT systems are a fusion of IT and OT [4], cybersecurity issues related to IT are encompassed by issues in IoT systems. An example is shown below.

- Implement countermeasures for known vulnerabilities
- Consideration of countermeasures based on threat analysis and risk assessment
- Avoidance and detection of bugs and lack of functionality during implementation
- Communication security (FW, IPS/IDS, encryption, etc.)
- Implementing authentication and access control
- Protect and update credentials and trust anchors
- Logging and analysis

The research on securing IT systems that has been conducted over the years has accumulated results and can be utilized in IoT systems, and thus is excluded from consideration in this paper, except for "authentication and access control," which is discussed below [4]. In cases where utilization is difficult due to limitations of individual IoT devices (e.g., computing resources), it is necessary to conduct a separate study. On the other hand, this paper will focus on the following items [4], which are listed as characteristics of IoT systems.

a) The scope and impact of the threat can be very broad and large.
b) The demand for IoT systems, especially in operation and maintenance, is more than 10 years.
c) Monitoring and managing IoT devices can be very difficult. There may be unmanaged IoT devices.
d) IoT devices may have difficulty being fully aware of each other's environment
e) IoT devices have limited functionality and performance.
f) Potential connection of IoT systems not envisioned by the developer

These characteristics raise the following concerns: b) vulnerabilities discovered after the IoT device is shipped may affect the response to vulnerabilities; c) vulnerabilities may be left unpatched and un-updated; e) patching may be difficult due to IoT device resources; f) the IoT device may be vulnerable to security risks; and g) the IoT device may be vulnerable to security risks. e), there may be situations where it is difficult to apply patches to IoT devices due to resource constraints. With regard to the provision of these patches and updates, the following points have been raised from an economic point of view [9].

- Currently, device vendors and manufacturers have little financial incentive to ensure continuous upgrading of patches for IoT
- Maintenance of IoT devices can reduce revenues because company revenues come from device sales, not maintenance
- Vendors are not legally responsible for the ongoing maintenance of the device after the initial sale (Japan's Ministry of Land, Infrastructure, Transport and Tourism ensures safety through its vehicle inspection system [10].)
- Vendors tend to pursue planned obsolescence of devices to maximize profits through continued sales rather than maintaining existing devices
- Bankruptcy or dissolution of the company or organization that managed the IoT device

The risks associated with these conditions, characteristics, and point out need to be considered for IoT systems, including IoT actuators.

4 DIRECTION OF SECURITY MEASURES IN IOT ACTUATORS

4.1 Output Access Control to Physical Space

IoT systems are diverse, and it has been noted that even when the IoT devices used and system configurations are similar, the security measures required differ depending on the purpose and application of the IoT system [11]. Prior research on security countermeasures for output to physical space has focused on threats to loudspeakers, which are considered one of the IoT actuators. In that research, as a countermeasure against analog signal threats to sensors, the Cyber-Physical Firewall (CPFW) framework, which detects and regulates the analog signal output of the CPS using a mechanism similar to a firewall, has been proposed [12]. It has been shown that CPFW is effective as a countermeasure against threats posed by sound signals by controlling output access to the digital signal output of the controller with CPFW, and outputting only secure digital signal output from the output interface after digital-to-analog conversion. For sound signals, the report also cites the wide variety of threats presented among analog signals and formulates access control policies for known attacks as a case study evaluation.

On the other hand, the report states that, unlike loudspeakers, the case of motors and other devices that cause changes in the environment due to motion is considered a future issue, and furthermore, secondary effects resulting from actuators require a mechanical engineering approach.

With reference to the above, it is considered effective for security measures in IoT actuators to introduce policy-based access control mechanisms in the course of signal transmission between the controller and the output interface. If the policy formulation is to be specific to individual threats, consideration should be given to increasing the number of policies and ensuring real-time performance.

In addition, the relationship between the information (ATTRIBUTE) that can be extracted from the analog signal and the threat signal is considered to differ depending on the type of actuator signal other than the sound signal, and this relationship needs to be clarified.

Furthermore, even if access control of output interfaces could regulate threat signals, there is concern that secondary effects could lead to more serious situations, so it is also

necessary to ensure consistency with high-level policies in terms of service continuity for IoT systems.

4.2 Continuous Access Control based on Dynamic Policies

Today's IT systems operate in complex network environments, including integration with cloud services and the use of terminals in mobile environments for remote work. Such complexity makes it difficult to define a single boundary between the inside and outside of an organization, leading to the compromise of conventional perimeter defense security measures. Therefore, a new concept known as "Zero Trust" is currently attracting attention, and a cybersecurity architecture called "Zero Trust Architecture" based on it [7]

4.2.1. Zero Trust, Zero Trust Architecture

Zero Trust focuses on the protection of resources and is based on the premise that trust is never implicitly granted, but must be continually assessed. This is because it assumes that attackers exist inside the perimeter and are just as untrustworthy as those outside. It is also assumed that resources are not limited to data, but include computational resources and IoT actuators. And to reduce cyber attacks by attackers that are difficult to eliminate completely, the focus is on authentication, authorization, and reduction of implicit trust zones to maintain availability and minimize the time delay of authentication mechanisms. Access control rules are supposed to be as granular as possible in order to enforce the minimum privileges necessary to execute a request. A conceptual diagram of zero-trust access is shown in Figure 1.

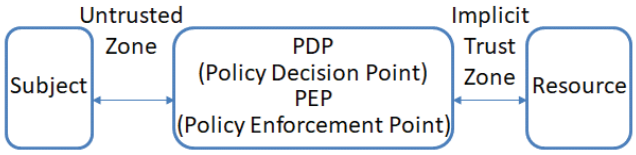


Figure 1: Conceptual Diagram of Zero Trust Access

When a subject accesses a resource, it will go through a PDP (Policy Decision Point) and a corresponding PEP (Policy Enforcement Point). However, no additional policies can be applied beyond the PDP/PEP. Therefore, the implicit trust zone should be as small as possible. The basic tenets of Zero Trust are as follows.

1. Consider all data sources and computing services as resources
2. Protects all communications regardless of network location
3. Access to resources is granted on a per session basis
4. Access to resources is determined by a dynamic policy that includes client ID, application service, state of the requesting asset, and other behavioral and environmental attributes
5. Monitor and measure the integrity and security behavior of all assets
6. Dynamic authentication and authorization of all resources, strictly enforced before access is granted
7. Gather as much information as possible about the current state of assets, network infrastructure, and communications and use it to improve security posture

Behavioral attributes, shown in 4. above, include analysis of the accessing subject and deviations from observed usage patterns. Environmental attributes include the timing of access and detection of currently active attacks. Since these attributes are affected by changes in time, history, and circumstances, access is determined by dynamically applying a policy, a set of access rules that take them into account.

Figure 2 is a conceptual framework model showing the logical components that comprise a zero-trust architecture and the basic relationships among their interactions.

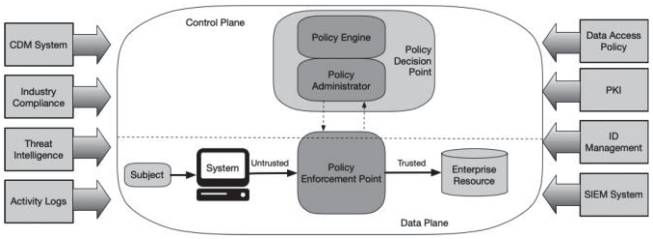


Figure 2: Core Zero Trust Logical Components [7]

In the above figure, the PDP is decomposed into two logical components: the PE (policy engine) and the PA (policy administrator).

- PE (Policy Engine)

The final decision to grant access to a subject's resources is made by the Trust Algorithm (TA). Use external sources (CDM System, Threat Intelligence, etc.) and policies as input to the Trust Algorithm to control access to resources.

In the United States, NIST is developing the Policy Machine [13] as a reference implementation of NGAC (Next Generation Access Control) [14], a standard for ABAC (attribute-based access control) based on relationships among data elements, with the goal of supporting a wide variety of access control policies [15].

- PA (Policy Administrator)

The PA works closely with the PE and relies on the PE to decide whether to allow or deny access; the PA causes the PEP to establish or block communication paths between entities and resources based on the PE's decision.

- PEP (Policy Enforcement Point)

The PEP works with the PA and receives policy updates from the PA. It then applies the received policy to establish, block, and monitor the communication path between the subject and the resource.

- TA (Trust Algorithm)

The thought process in determining access permissions in PE is considered TA. This thought process is grouped into the following categories.

A) Access request
B) Subject database and history
C) Asset database
D) Resource policy requirements
E) Threat intelligence and logging

In addition, TA implementation methods are distinguished by two main features.

1. Criteria-based (binary)/score-based (weighted)
2. Unitary (individual request) / Contextual (history consideration)

Contextual score-based TA is thought to provide dynamic and fine-grained access control.

5 APPLICATION OF ZERO TRUST ARCHITECTURE TO IOT ACTUATORS

In an IoT system, IoT actuators provide various added values by cooperating with IoT services on the cloud. In this case, the application running on the IoT actuator controls the actuator in the physical space in response to requests from the IoT service. Note that the IoT actuator may have another transducer function, the sensing function. A schematic of an IoT actuator and IoT system is shown in Figure 3.

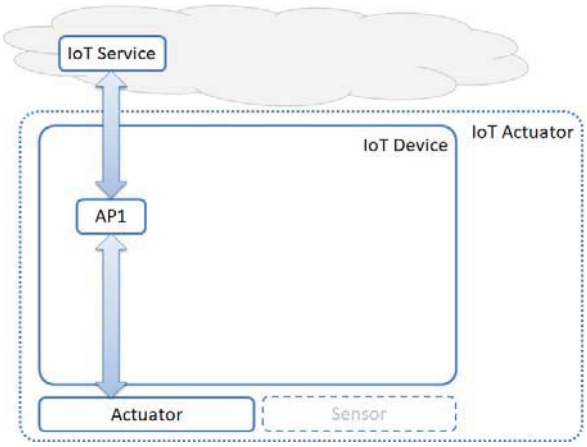


Figure 3: Schematic of IoT actuator and IoT system

One of the most common types of actuators in the physical space is the motor. A motor is a device that uses electric power to generate physical rotational motion. This rotational motion can be used to drive a pump in a river disaster prevention system or a water supply system for paddy fields, or to rotate the propeller of a drone or the wheels of an automatic delivery robot. Figure 4 shows an example of the device configuration of the IoT actuator with motor envisioned in this paper [16].

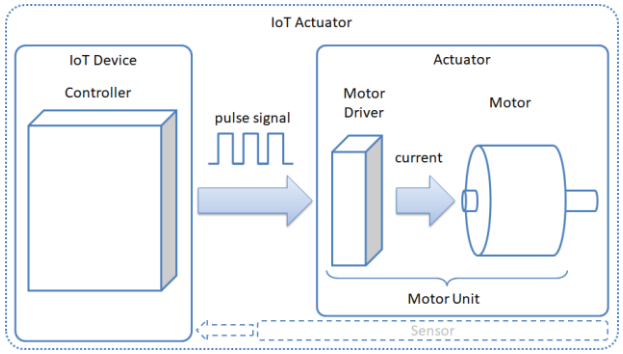


Figure 4: Example of a possible IoT actuator device configuration

As shown in the figure above, an IoT device corresponds to a controller. A motor unit consisting of a motor driver and a motor corresponds to an actuator. As a security measure for IoT actuators, it is considered effective to introduce a policy-based access control mechanism during the signal transmission between the controller and the output interface. Therefore, we will consider implementing a function equivalent to CPFW in IoT devices.

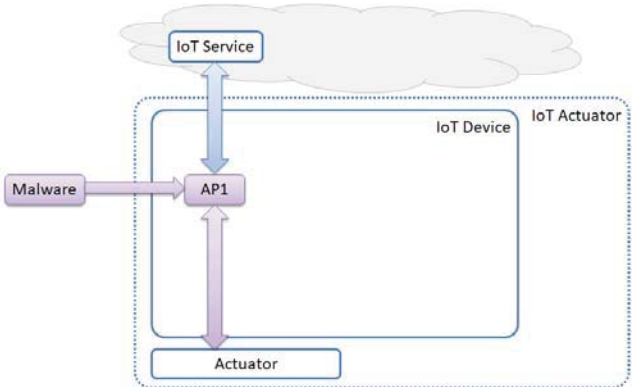


Figure 5: Example of a cyber attack against an IoT actuator

Malware that attacks IoT actuators is assumed to infect applications (software) running on IoT devices and launch cyberattacks against actuators. However, as pointed out in Section 3, it is not always possible to update the vulnerable programs as soon as possible. However, as pointed out in Section 3, there are cases where the provision of programs and patches for updates are delayed or not provided. Furthermore, in recent years, zero-day attacks, in which software vulnerabilities are discovered and exploited before patches or workarounds are released, have become a serious problem [17].

Therefore, in order to prevent malware-infected applications from controlling unauthorized actuators, a function equivalent to CPFW is considered based on the zero-trust concept. First, PDP and PEP in in the zero-trust architecture, PDP is placed on the cloud side and PEP is placed on the IoT actuator side. This PEP will act as a CPFW to prevent unauthorized control. In addition, the PDP can work with IoT services to obtain behavioral and environmental attributes of the subject application. This

allows the PEP to work with the PDP to provide instrumental access control based on dynamic policies. A schematic of the application of Zero Trust Architecture in an IoT actuator is shown in Figure 6.

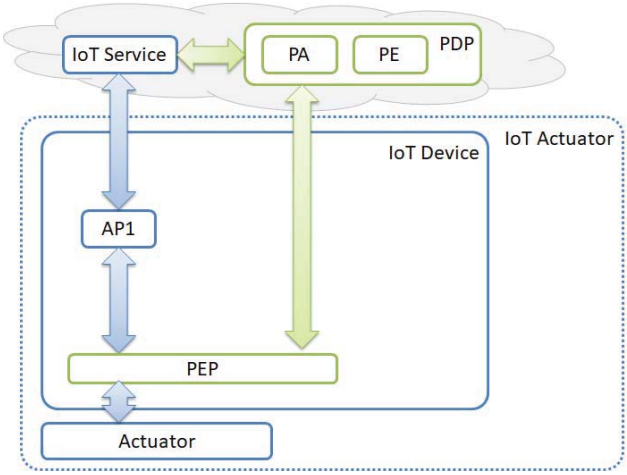


Figure 6: Application of ZTA in IoT actuator

However, there is concern that PEPs added to IoT actuators will become a new target for malware attacks: even if PDPs are protected by cloud-side security measures, without proper access controls being enforced by PEPs, attacks on the actuators could be carried out and damage to the physical space.

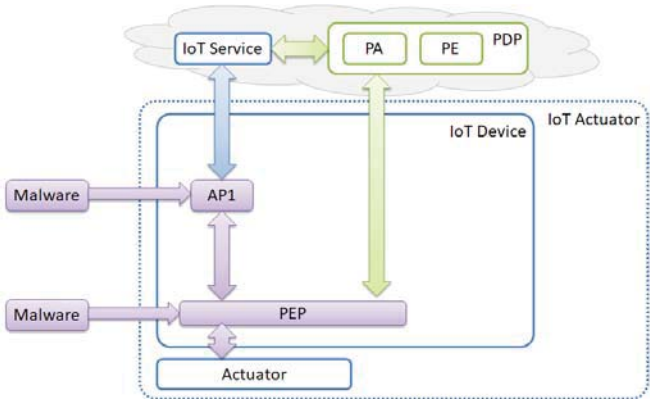


Figure 7: Example of a cyber attack against PEP

Therefore, we focus on TEE (Trusted Execution Environment) [18], which is a hardware security feature of SoC (System on Chip), a kind of semiconductor chip that is being adopted in IoT devices such as IoT actuators.

TEE is an isolated execution environment independent of apps and OSs, and the Secure World in the isolated execution environment is not affected by the Normal World in which apps and OSs are executed. Therefore, PEP can be placed inside Secure World to protect it from malware and other attacks. Examples of hardware that provide TEE include Arm TrustZone [19], RISC-V Keystone [20], Intel SGX [21][22], etc. A schematic of PEP protection by Secure World is shown in Figure 8.

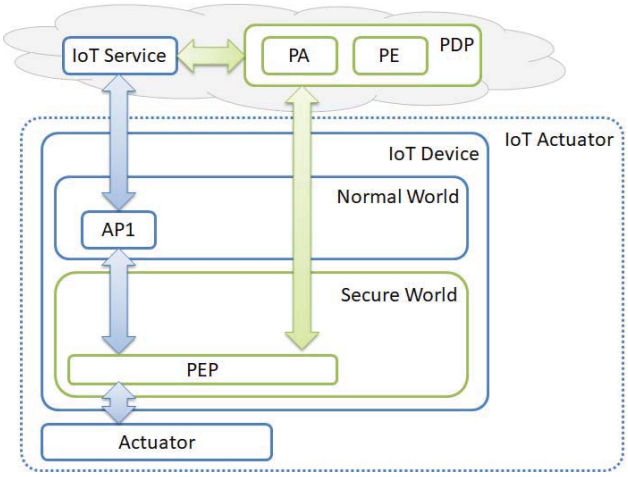


Figure 8: Protecting PEP with Secure World

The above configuration makes it possible to prevent unauthorized control of actuators by applications in which the PEP is infected with malware. It is also possible to prevent the PEP itself from being infected with malware. However, there is a risk of a denial of service attack (DoS attack) in which a malware-completed app interferes with the coordination between the PEP and PDP. In addition, there is a risk that IoT actuators will not provide the services they are supposed to fulfill because normal apps will not control the actuators as they should.

Figure 9 shows an example of a DoS attack on an IoT actuator. Since IoT actuators are expected to be used in applications where the availability of object motion is important, countermeasures against such DoS attacks are necessary.

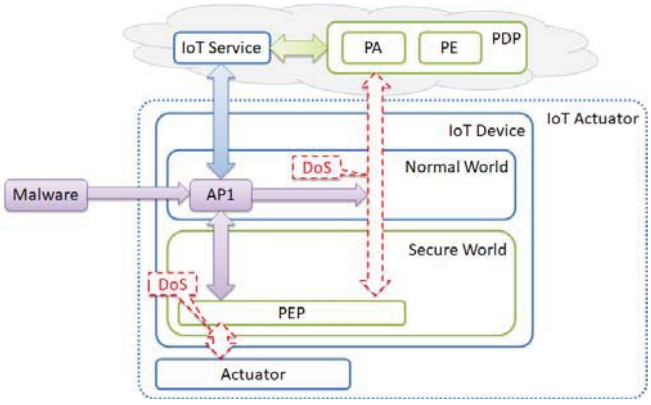


Figure 9: Example of cyber attack causing DoS

Accordingly, Self PDP, which is a subset of PDP, is introduced to the IoT actuator in preparation for the situation where the PDP on the cloud side and the PEP in the IoT actuator cannot be linked. However, since the input to the PE of Self PDP is limited to within the IoT actuator, the process by the TA is limited. On the other hand, the communication resources required for the linkage between PEP and Self PDP are no longer necessary, and real-time performance and availability are expected to be improved.

In addition, in order to mitigate secondary damage to IoT actuators caused by normal applications not controlling the

actuators as they should, applications with Failsafe functionality will be introduced to IoT actuators. For example, provide a function to control the actuator to stop safely while it is running, such as by gradually decreasing the motor speed. In addition, as an application of the Failsafe function, it is also possible to realize EOL (End of Life) for IoT devices. The management of EOL by the Failsafe function is an advantage in preventing unmanaged IoT devices from being abused by DDoS attack bots, etc., which are related to the characteristics b) and c) of the IoT system described in Section 3. In addition, as a countermeasure against malware attacks, Self PDP and Failsafe AP must be placed in the Secure World in the same manner as PEP.

As a result of the final consideration, the relationship between the components that apply the Zero Trust architecture to IoT actuators is shown in Figure 10 as the “Zero Trust IoT Security Framework”.

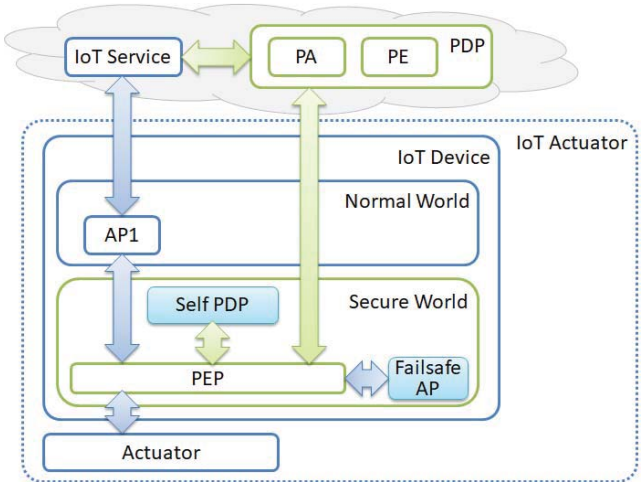


Figure 10: Zero Trust IoT Security Framework

6 CONCLUSION

This paper discussed, in particular, the results of a study of security in IoT actuators that exert physical action as feedback from cyberspace to physical space. The paper also presents the direction of future security measures in IoT actuators. Then, he proposed the "Zero Trust IoT Security Framework," which applies the Zero Trust architecture to IoT actuators. In the future, we plan to work on the detailed design of the "Zero Trust IoT Security Framework", implement it, and verify its effectiveness.

REFERENCES

[1] "Next Cyber Security Strategy (Draft)", Cabinet Cyber Security Strategy Headquarters, September 27, 2021, https://www.nisc.go.jp/conference/cs/dai31/pdf/31shiryou01.pdf

[2] Zen Ishikura, Hideo Yamamoto, Takemi Nisase, Fumihiko Magata, "Considerations on Security Design of IoT Systems," 2018 Symposium on Cryptography and Information Security (SCIS 2018), 4E2-1, January 2018

[3] RON ROSS, MICHAEL McEVILLEY, JANETCARRIER OREN, "NIST Special Publication 800-160 Systems Security Engineering," NIST National Institute of Standards and Technology U.S. Department of Commerce, November, 2016, https://doi.org/10.6028/NIST.SP.800-160v1

[4] Katie Boeckl, Michael Fagan, William Fisher, Naomi Lefkovitz, Katerina N. Megas, Ellen Nadeau, Danna Gabel O'Rourke, Ben Piccarreta, Karen Scarfone, "NISTIR 8228 Considerations for Managing Internet of Things (IoT) Cybersecurity and Privacy Risks," NIST National Institute of Standards and Technology U.S. Department of Commerce, June, 2019, https://doi.org/10.6028/NIST.IR.8228

[5] Michael Fagan, Katerina N. Megas, Karen Scarfone, Matthew Smith, "NISTIR 8259 Foundational Cybersecurity Activities for IoT Device Manufacturers ", NIST National Institute of Standards and Technology U.S. Department of Commerce, May, 2020, https://doi.org/10.6028/NIST.IR.8259

[6] Michael Fagan, Katerina N. Megas, Karen Scarfone, Matthew Smith, "NISTIR 8259A IoT Device Manufacturers Cybersecurity Capability Core Baseline, NIST National Institute of Standards and Technology U.S. Department of Commerce, May, 2020, https://doi.org/10.6028/NIST.IR.8259A

[7] Scott Rose, Oliver Borchert, Stu Mitchell, Sean Connelly, "NIST Special Publication 800-207 Zero Trust Architecture," NIST National Institute of Standards and Technology U.S. Department of Commerce, August, 2020, https://doi.org/10.6028/NIST.SP.800-207

[8] "White Paper on Information and Communications, 2020 Edition," Ministry of Internal Affairs and Communications, August 2020. https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r02/pdf/index.html

[9] George Corser, Glenn A. Fink, Mohammed Aledhari, Jared Bielby, Rajesh Nighot, Sukanya Mandal, Nagender Aneja, Chris Hrivnak, Lucian Cristache, Internet of Things (IOT) Security Best Practices: whitepaper, IEEE, 2017, https://standards.ieee.org/wp-content/uploads/import/documents/other/whitepaper-internet-of-things-2017-dh-v1.pdf

[10] [Attachment 1] Permission System for Specified Modification, etc. of Vehicles (Outline), Ministry of Land, Infrastructure, Transport and Tourism, August 2020, https://www.mlit.go.jp/report/press/content/001358067.pdf

[11] IoT Security Guideline ver 1.0, IoT Promotion Consortium, Ministry of Internal Affairs and Communications, Ministry of Economy, Trade and Industry, July, 2016, https://www.soumu.go.jp/main_content/000428393.pdf

[12] R. Iijima , T. Takehisa, and T. Mori, "Proposal and Verification of a Security Framework to Detect and

Regulate Threats by Analog Signals," Proc. of Computer Security Symposium 2021, pp. 79 - 86, Oct. 19, 2021.
[13] David Ferraiolo, Serban Gavrila, Wayne Jansen, "NISTIR 7987 Revision 1 Policy Machine: Features, Architecture, and Specification," NIST National Institute of Standards and Technology U.S. Department of Commerce, Oct. 2015, https://doi.org/10.6028/NIST.IR.7987r1
[14] INCITS 565-2020 - Information technology - Next Generation Access Control, Information Technology Industry Council, <ref> August 19, 2022, [15] https://standards.incits.org/apps/group_public/project/details.php?project_id=2328
[15] Policy Machine Core, GitHub, https://github.com/PM-Master/policy-machine-core
[16] Overview and Features of Stepping Motors, Oriental Motor Co, https://www.orientalmotor.co.jp/products/stepping/overview_1/
[17] Ten Major Information Security Threats 2022, IPA, July 2022, https://www.ipa.go.jp/files/000096258.pdf
[18] Ariyasu Suzaki, Akira Tsukamoto, Kazumoto Kojima, Kenta Nakajima, Hongtron Tuk, Akira Shiio: TEE Comparison, Symposium on Cryptography and Information Security (SCIS) 2020 (2019).
[19] S. Pinto and N. Santos, “Demystifying Arm TrustZone: A Comprehensive Survey,” ACM Computing Surveys (CSUR), vol. 51, no. 6, 2019
[20] D. Lee, D. Kohlbrenner, S. Shinde, D. Song and K. Asanovi´c: “Keystone: A framework for architecting tees”, arXiv (2019).
[21] V. Costan, I. Lebedev, and S. Devadas, “Secure Processors Part I: Background, Taxonomy for Secure Enclaves and Intel SGX architecture,” Foundations and Trends in Electronic Design Automation, vol. 11, no. 1-2, pp. 1-248, 2017.
[22] V. Costan, I. Lebedev, and S. Devadas, “Secure Processors Part II: Intel SGX Security Analysis and MIT Sanctum Architecture,” Foundations and Trends in Electronic Design Automation, vol. 11, no. 3, pp. 249-361, 2017.

Implementation on the Evaluation Platform for a Large-Scale Building Automation and Control System

Kohei Miyazawa*, Tetsuya Yokotani**, Hiroaki Mukai***

* Electrical and Electronic Science Engineering, Kanazawa Institute of Technology, Japan
c6100786@planet.kanazawa-it.ac.jp
**Engineering department, Kanazawa Institute of Technology, Japan
yokotani@neptune.kanazawa-it.ac.jp
*** Engineering department, Kanazawa Institute of Technology, Japan
mukai.hiroaki@neptune.kanazawa-it.ac.jp

Abstract—In this paper, we propose an evaluation platform based on a building automation and control system (BACS) using a building automation control network (BACnet) protocol. A BACnet connects electrical equipment with a monitoring center to manage and control multiple electrical devices in a building centrally. Although a BACnet has recently been widely deployed as a consensus solution, some security concerns have been identified for implementation expanded to the IP base protocol, that is, BACnet/IP. For example, unauthorized remote access to control systems for electrical equipment and denial of service attacks may lead to serious accidents, particularly in large buildings. Therefore, security threats are typically detected by monitoring traffic on BACnet/IP without complicated functions on the network endpoints. However, investigating the detection mechanisms in real building systems is relatively difficult. Therefore, we propose an emulated evaluation platform for personal computer (PC) systems. In this study, we clarify the requirements of a BACS, including the latest trends in the use of the platform in buildings, and describe an implementation of the proposed platform by PC clustering, which supports up to 100,000 endpoints. Diagram of the configuration of the evaluation platform was designed. Emulators were distributed at each level of the hierarchy to unify the platform. The monitoring center reads the status of all locations in one-minute cycles to analyze the network traffic at the maximum load. Finally, we show that the proposed platform can verify the performance with the maximum network load without using a real building environment.

Keywords: BACnet, Building automation and control system (BACS), Cyber physical system, Cyber physical security

1. Introduction

In recent years, communication protocols for building management systems have become more open. particularly a communication protocol called the building automation and control system (BACS), which became an ISO standard in 2003 as an international data communication protocol, has been applied to the automatic control of equipment in buildings. It has been applied to the automatic control of facilities in buildings. It is connected to various facilities, such as air conditioning, electric power, lighting, crime prevention, and disaster prevention, contributing to a comfortable environment for users. Additionally, as it can automatically and centrally manage equipment from different vendors, it contributes to improved convenience and safety.

As the demand for solar power generation and other electric power facilities increases to achieve higher energy efficiency in the future, the number of DC power feeders adopted is expected to rise, and equipment related to these facilities will also be controlled via a building automation control network (BACnet) [1]. However, when a BACnet is designed, it is always recommended that BACnet devices be installed in a separate segmented network. With the recent rise in the Internet of things (IoT), building automation system (BAS) networks are now connected to the Internet for management coordination [2]. Interconnectivity between networks facilitates attacks between networks. When DC power supply equipment, which often handles high voltages, is attacked in such cases, it is suggested that arcing may occur in the switchgear, leading to a fire accident, and the damage if such an incident occurs in a large building is immeasurable. However, currently, there is little awareness of the threat posed by external attacks on DC power feeders. Moreover, BACnet is used in small- and medium-sized buildings, commercial facilities, airport systems, and large buildings to perform control and management tasks in a single package, thereby enabling long-term operations. As previously mentioned, it is necessary to evaluate the impact of an attack. However, if the evaluation is conducted in an actual environment using the electrical facilities of a large-scale building, it is necessary to install electrical facilities that are adapted to the assumed building environment, which can be costly [3].

Thus far, the authors have focused their research only on the physical impact of external unauthorized attacks via a BACnet [4]. In this study, the previous evaluation platform was modified from AC to DC, and the temperature characteristics of the switchgear under a DC power supply were verified. Based on the results, we design and emulate the logical configuration of a large-scale building with 100,000 monitored points to respond to identified threats. The emulated building is distributed at each level of the hierarchy to unify the platform, and the network traffic under the maximum load environment is investigated and analyzed.

2. Overview of a BACnet/IP

A BACnet is a communication standard for building networks that became an ISO standard in 1995, similar to ANSI ASHRAE Standard 135-1995. The BACnet Operator Workstation (B-OWS), a central monitoring device, interconnects building facilities through a common interface

called the BACnet device to monitor and control the building [5].

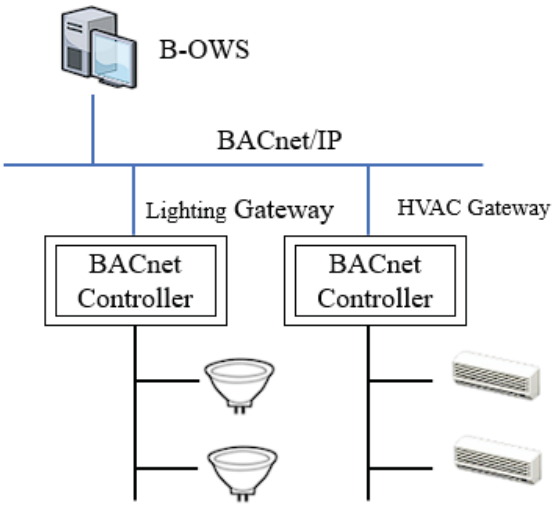


Fig.1 BAS configuration

A BACnet/IP is a technology in which the upper protocols of the BACnet standard are implemented on an IP network, as shown in Fig.1. This implies that various BACnet-enabled devices are connected to the Ethernet, which is commonly used in all facilities. By using existing networks, interoperability between systems and the cost of configuring network infrastructure can be reduced. The Category 5 cabling that a BACnet/IP uses for communications is available in virtually all areas of the building. All of these factors reduce costs in terms of initial deployment and ongoing maintenance and management of the BAS network.

However, it is necessary to understand the problems associated with a BACnet/IP. One particularly critical issue is security: a BACnet/IP is a standardized, interoperable network and is therefore not suitable for use on unprotected networks. Therefore, the defense of building management system networks with BACnet/IP in IP networks is critical.

2.1 BACnet object

A BACnet represents the target data as objects, which is an abstract concept, and models the facilities connected to the network as a collection of objects, which are classified into basic input/output, device characteristics, notification functions, life and safety, complex functions, file information exchange, and others. BACnet objects are classified into basic input/output, device characteristics, notification function, life and life safety, complex functions, file information exchange, and others. The ASHRE 135-2012 BACnet standard specifies 54 types of BACnet objects. Each object defines several attribute values, called properties, to further define the characteristics of that object in detail [6][7]. Table 1 shows the typical objects in the BAS handled in this study: analog input and analog output are used to monitor analog value information, such as room temperature and set temperature (the former for monitoring, and the latter for control). Multi-state input and output are used to control and monitor multiple states, such as the operating mode and airflow rate.

An accumulator is used to monitor the total amount of electricity. These objects have a present-value property, and changes in that value allow the device to physically control and monitor its state. As these objects have a physical effect on the real environment when abnormal values are assigned to them, it is necessary to be aware that an unauthorized attack can pose a serious threat.

Table.1 Typical BACnet Objects

Object	Property
Analog Input	Present Value
Analog Output	Present Value
Binary Input	Present Value
Binary Output	Present Value
Multi-state Input	Present Value
Multi-state Output	Present Value
Accumulator	Present Value

2.2 BACnet service

Access to objects owned by BACnet devices is provided by service requests. Various services are implemented, and the main ones used are those for creating and deleting objects (created object, deleted object), reading and writing properties individually, and multiple properties at once (read property, write property), and receiving and sending change notifications (change of value). There are equally services for receiving and sending change notifications (change in value), etc. [8].

The number of devices and objects increases as the number of buildings increases. When monitoring the status of controllers at regular intervals, the amount of data to be read increases, and the monitoring equipment may become overloaded.

3. Evaluation of the physical phenomena

Many electrical devices in building facilities operate using direct current (DC) power. However, current power distribution facilities transmit power in AC, which must be converted to DC when an electrical equipment is in operation, resulting in energy loss. Eliminating the need to convert DC to AC reduces energy loss. As an IT equipment that requires a continuous power supply becomes more established, the amount of power consumed increases, and power consumption needs to be reduced [9]. To solve these issues, energy savings by DC power supply are being promoted. In line with this trend, an increasing number of energy-related equipment, such as photovoltaic power generation equipment, is being incorporated as a part of construction facilities. Consequently, the amount of photovoltaic power generation installed has increased remarkably over the past 9 years. Fig.2 shows the cumulative power generation from renewable energy sources in Japan (excluding large-scale hydropower) [10]. The installed capacity of solar power generation has increased by approximately eightfold to 74.7 million kW over 9 years from FY2013 to FY2021. This indicates that the demand for the DC power feed is high and is expected to increase further in the future.

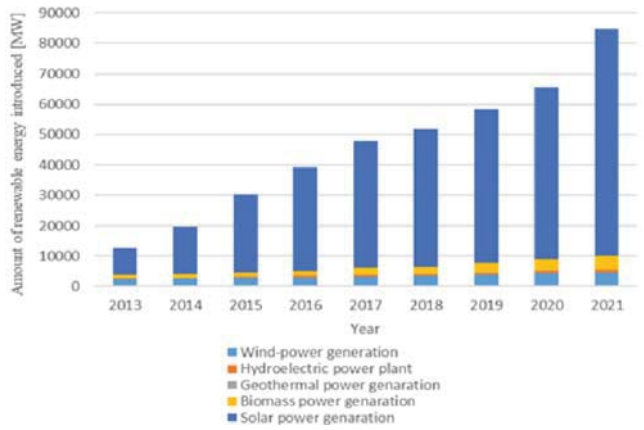


Fig.2 Cumulative power generation from renewable energy facilities

3.1 Hazards of DC power supply

Fig.3 shows the fire process for a hypothetical DC feed [11].

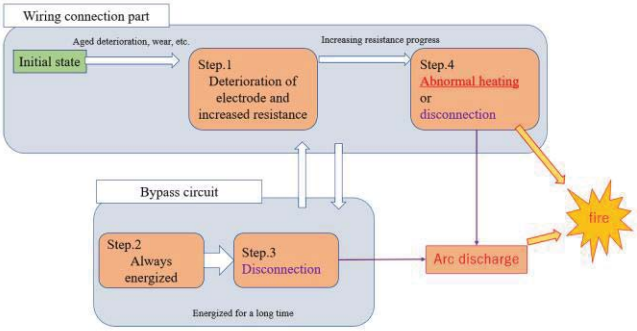


Fig.3 Fire process for DC feed

Although DC power supply has many advantages and is in high demand, it also has disadvantages. Generally, arcing occurs when power is interrupted in both DC and AC; however, in the case of AC, arcing stops spontaneously because there is a point when the voltage reaches zero. However, with direct current, the voltage is constant; as long as power continues to be supplied, arcing will not stop spontaneously. The arc generates heat, which can cause welding between the contacts and possibly cause a fire. Furthermore, in the case of photovoltaic power generation, the system is characterized by the fact that it continues to generate power as long as it is exposed to sunlight. Therefore, once an arc is generated in a photovoltaic power generation system, it will not disappear spontaneously but will continue to generate power, and there is a risk of damage escalation. This also implies that the system will continue to generate power in the event of a fire, which is dangerous from the perspective of firefighter protection.

3.2 Example of BACnet use with DC power supply

This is an example of a BACnet use in solar power generation. Currently, the DC electricity generated by solar panels is converted to AC by a power conditioner before usage. However, as the shift to DC progresses in the future, the electricity generated by solar panels will be used as direct current, and the electrical equipment at the end of the line will be controlled via BACnet.

4. DC-Fed BACnet evaluator

In many existing buildings, electrical equipment is operated by AC power transmission, and the evaluation platform created last year was designed to evaluate the temperature rise characteristics by operating the electrical equipment with an AC power supply. However, as mentioned above, it is expected that DC will be used to supply power to electrical equipment in the future instead of inefficient AC. To accommodate this, the evaluation platform of the previous year should be improved.

The BACnet controller is implemented using a programmable logic controller and synchronized with the switchgear, which is operated by sending high-frequency on/off signals to the switchgear at regular intervals via BACnet communication from the B-OWS. In this case, the short-circuit and temperature rise characteristics owing to the welding of the contact points when an arc discharge occurs in the switchgear operated by the DC power supply are evaluated.

4.1 Design and improvement of a BACnet evaluation machine

Fig.4 shows the configuration of the BACnet demonstration unit designed and modified in this study. The areas shown in red were changed to support the DC power supply. Figs.5 and 6 show the transformer, rectifier, and resistors.

The existing system was designed to measure high-frequency cutoff temperature characteristics at low AC voltages, and it was modified for testing at 280 V DC. The following areas were focused on when improving the existing demonstrator:

a. Selecting electromagnetic switches and breakers compatible with high voltages and replacing them with existing parts.

b. Modifying the electromagnetic switchgear by installing a transformer to allow the electromagnetic switchgear to draw approximately 280 V.

c. Installing a rectifier and replacing the existing parts with new ones.

d. Providing a rectifier to allow the electromagnetic switch to draw DC power.

e. Installing breakers on the primary and secondary sides to prevent overcurrent, leakage currents, etc. because large currents are expected to flow.

f. Two exhaust fans are used to exhaust the generated heat. Because the heat generation is expected to be large, the fans have been modified to become more powerful.

g. Large resistors are installed to handle high voltages.

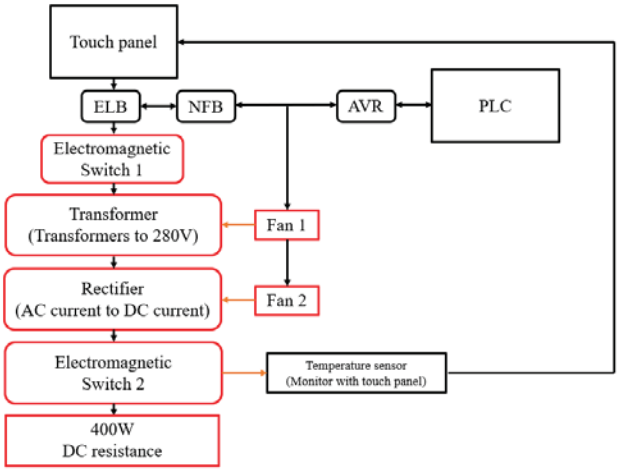


Fig.4 Configuration of BACnet evaluation machine

Fig. 5 Power transfer

Fig.6 Rectifier

4.2 Experiment configuration

This experiment assumes a scenario in which an electrical facility is under attack from the outside via a BACnet during the night when the administrator is asleep, which is turned on and off repeatedly at a high speed from the BACnet device for a long period to evaluate the appearance of the switchgear and the characteristics of the temperature rise near the switchgear. Temperatures were measured at five locations: outside air temperature, at the switch connector, left side of the switch, right side of the switch, and temperature inside the panel.

Fig.7 shows the locations of the temperature sensors and the location inside the panel.

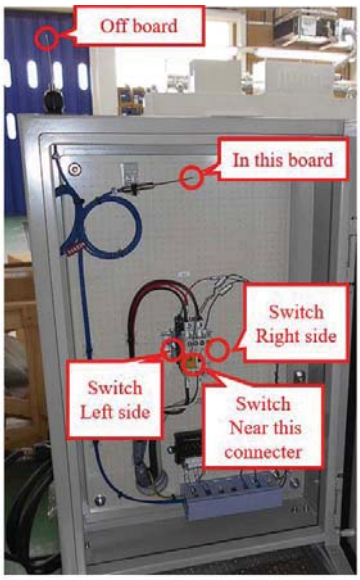


Fig.7 Temperature sensor position of evaluation machine

4.3 Experiment summary

Fig.8 shows the switchgear issues due to arc discharge.

Fig. 8 Arc discharge

Fig.9 shows the experimental results at a 6-h/on/off interval of 1 s.

Regarding the temperature rise characteristics, the connector part of the switchgear exhibited a steeper temperature rise than the other parts, and the maximum temperature was also the highest.

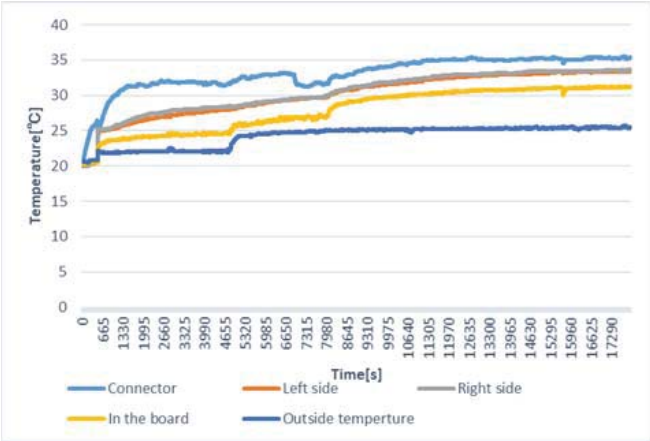


Fig. 9 Temperature characteristics of switchgear

4.4 Experimental results

BACnet evaluation equipment was designed and improved to support the DC power supply. The improved equipment was then operated to verify the temperature characteristics of the switchgear. In this experiment, the switchgear did not reach temperatures that could cause ignition; however, arcs caused by switching the switchgear on and off were confirmed. Therefore, although it is unlikely to be a direct cause of ignition, the degradation of the switchgear owing to arcing poses a certain threat during long-term operations.

5. Evaluation platform on large-scale building systems

The certainty of the attack was demonstrated in the previous section, as the threat of the actual physical impact was extracted from the heat rise caused by the continuous operation of the switchgear owing to the DC power feed. Such an attack on each BACnet device in a large building would cause rapid traffic in the network between them. This significantly degrades the network performance and disrupts the actual environment.

Therefore, we propose and construct an environment to emulate a large-scale building with 100,000 monitoring points of electrical devices, or properties of objects possessed by BACnet devices, in this evaluation environment building, and generate burst traffic by reading all monitoring points in a 1-min cycle to evaluate the impact of the attack. Propose and construct.

5.1 Maximum configuration overview

The configuration of a large building with 100,000 monitored points, proposed to generate the maximum load, is shown in Fig.10.

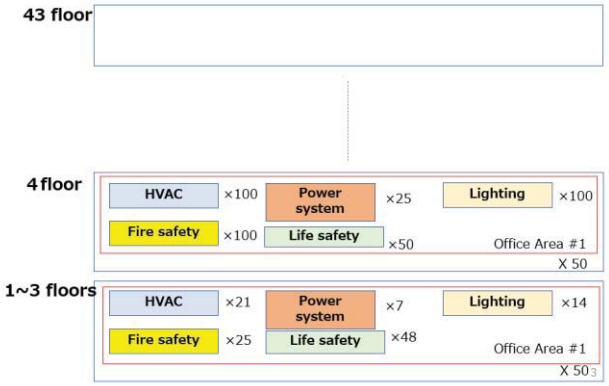


Fig.10 Proposed model of 100,000 monitored points

The model was created considering a building with 43 floors, as shown in Fig.10, with 50 rooms per floor and five controllers per room. Each controller was assigned an element to monitor the equipment in the building, and the number of elements varied depending on the role of the controller. Fig.11 shows the specific role of each object of the controller.

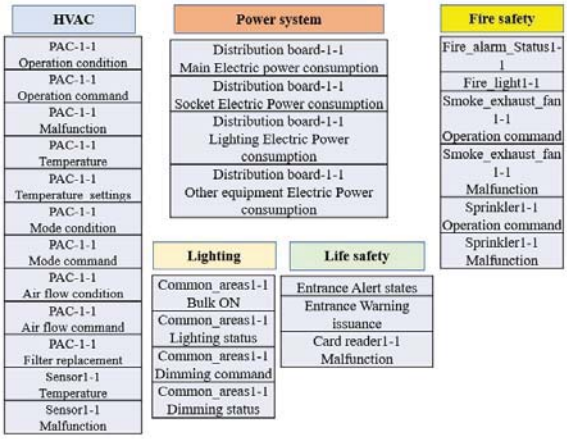


Fig.11 Attributes of each monitored object

Each of the five controllers was assigned 12 objects for air conditioning, 4 for power, 4 for lighting, 6 for fire prevention, and 3 for security, each of which is used to monitor and control the facilities. The number of monitored objects per floor was multiplied by the number of objects in each controller, multiplied by the number of floors.

5.2.1 Emulation on environment configuration

The emulation structure of the BACnet-style traffic-generation model is shown in Fig.12.

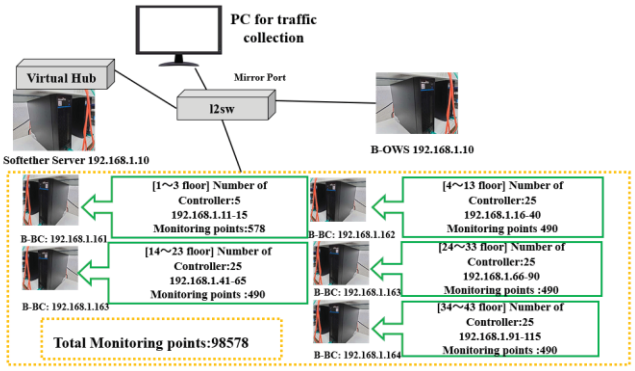


Fig. 12 Network configuration

It is difficult for a single PC to emulate 43 levels of controllers because of performance limitations. Therefore, as shown in Fig.12, each client computer emulated 15 controller levels and clustered them. To approximate the real environment, each controller is assigned a virtual NIC with an IP address (192.168.1.11-105) and a MAC address. They received control monitoring from the B-OWS via a switch. Additionally, a PC for traffic collection was installed to check the current status.

5.2.2 Connectivity by SoftEther VPN

To make it as close as possible to a real environment, it is necessary to assign a network interface card (NIC) to each BACnet device and allocate IP and MAC addresses to them. Therefore, this platform uses L2-VPN software called SoftEther VPN to generate and apply virtual NICs.

The role of the SoftEther VPN is to realize a VPN between locations to create a virtual private network [12]. As shown in Fig.12, a virtual HUB is created on a server computer with a SoftEther VPN installed, and it accepts VPN connections with client computers that have BACnet devices. NICs can be created, and the computer that creates the virtual NICs can establish VPN connections to the virtual HUBs created on the VPN server. Totally, 105 virtual NICs were assigned to client computers with 105 BACnet devices, each with a different IP and MAC address. Additionally, SoftEther VPN has a feature called a local bridge, which allows a bridge connection to be established between the virtual HUB of the VPN server and an existing physical Ethernet segment. Using this feature, it is possible to bridge B-OWS and the VPN server, and exchange packets with the virtual HUB to which the clients' PCs are connected, thus achieving network connectivity on the same segment.

5.5 BACnet emulator

The BACnet testing tool monitor (TTM), which is commercially available, was used to emulate B-OWS for building monitoring. Instead of the actual B-OWS, various BACnet services, such as ReadProperty and WriteProperty, are sent to the controller to monitor the building status. Fig.13 shows the BACnet TTM operation screen.

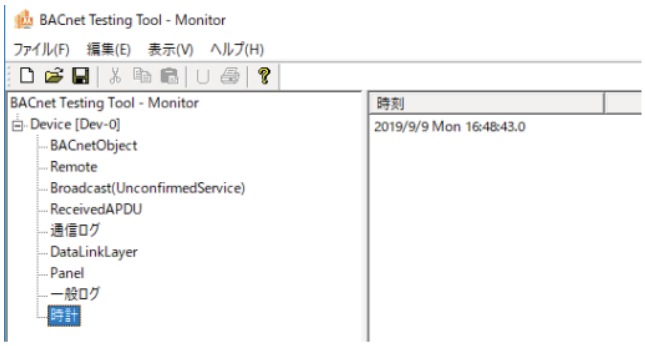


図 13 BACnet emulator

The Internet Explorer-like tree makes visual operations easy. Moreover, the tool's largest function is the "panel," as shown in Fig.14. All BACnet device properties registered in the panel can issue ReadProperty at the intervals set by the panel. That is, it is possible to read the monitoring points of the BACnet device at a certain cycle.

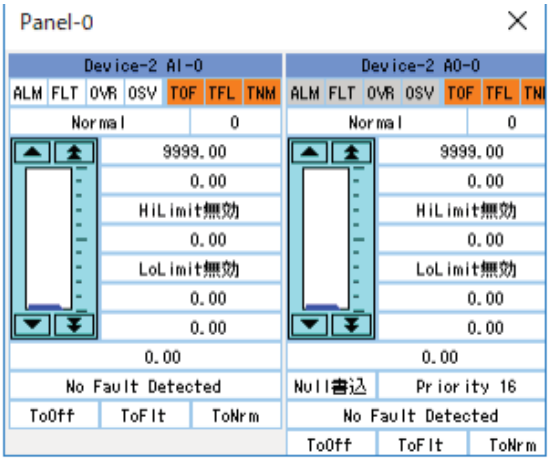


Fig.14 Operation panel

In this environment, this function is used to read in a virtual large-scale building with 100,000 monitored points. On the controller side, an emulator tool called ICONT Simulator is used to create a controller for each level of the hierarchy and execute various BACnet services.

5.4 Operation overview

This section describes the operational verification of the large-scale building emulation environment constructed based on the description in the previous section. This generates a pseudo traffic burst. The traffic between the B-OWS and client PC is then captured, and the communication effect is investigated. Fig.15 shows the panel loading screen.

Fig.15 Panel loading

Due to the restriction that only 50 properties can be opened per panel, it is necessary to open as many panels as the total number of monitored points divided by 50. Furthermore, if all monitored points are opened simultaneously, the software will stop working due to its specifications, so all 50 panels must be opened at 120 s intervals.

5.5 Operation results

Fig.16 shows the traffic transition when the model shown in Fig. 10 is continuously loaded in 60 s cycles.

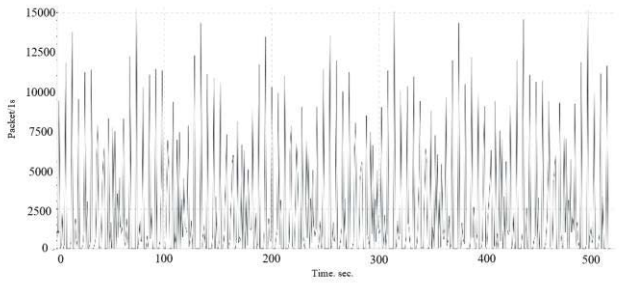


Fig.16 Read all monitored points

6. Conclusions and future works

In this study, we improved the BACnet evaluation system we developed in the previous paper. As a result of assuming that the high frequency of on/off signals is an unauthorized attack, we found that physical abnormalities occur in the switchgear in the DC power supply. We then proposed a maximum logical configuration that emulates a large-scale building and proposed an environment in which the amount of network traffic can be verified by reading it at regular cycles. In the future, we plan to develop a tool that can easily change the configuration of the maximum environment to enable flexible evaluation of network traffic according to the building environment.

Acknowledgments

Part of this work was supported by the Cabinet Office (CAO), Cross-ministerial Strategic Innovation Promotion Program (SIP), “Cyber-Physical Security for an IoT Society” (funding agency: NEDO). The authors sincerely thank the members concerned.

References

[1] Y. Takeuchi, “Current Situation and Future Prospects Among NVDC (Medium Voltage Direct Current) Distribution Network System” The Journal of the Institute of Electrical Installation Engineers of Japan, Vol.38 No.7 pp.365-368 (2018)

[2] O. Gasser, Q. Schitle, B. Rudolph, C. Denis, N. Schricker and G. Carle "The Amplification Threat Posed by Publicly Reachable BACnet Devices", Journal of Cyber Security and Mobility, Vol.6, No.1, pp.77-104 (2017)

[3] G. Stamatescu, I. Stamatescu, N. Arghira, and I. Făgărăşan “Cybersecurity Perspectives for Smart Building Automation Systems”, In Proceedings of the 12th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Bucharest, Romania, 25–27, pp.1–5 (2020)

[4] DG. Holmberg, “BACnet Wide Area Network Security Threat Assessment”, Nist Interagency/Internal Report (NISTIR) (2003)

[5] K. Miyazawa, T. Yokotani, and H. Mukai, “Development of a benchmark system on power-related control by BACnet/IP”, IWIN2021, Session8, No.28 (2021)

[6] ASHRAE.BACnet – A Data Communication Protocol for BuildingAutomation and Control Systems Addendum 135-2012aj (2016).

[7] K. Tomizawa, “Object and Property”, The Journal of the Institute of Electrical Installation Engineers of Japan, Vol.32, No.2, pp-115-118 (2012)

[8] M. Nast, B. Butzin, F. Golatowski, and D. Timmermann, “Performance analysis of a secured BACnet/IP network,” in 2019 15th IEEE International Workshop on Factory Communication Systems (WFCS), pp.1–8 (2019)

[9] AM. Fedel, AM. Ibrahim, AK. Tamazin, RA. Hamdy, and Ayman S. Abdel-Khalik, “IoT-Based Power Management for DC Microgrids”, 21st International Middle East Power Systems Conference (2019)

[10] Agency for Natural Resource and Energy, “Act on Special Measures concerning Promotion of Utilization of Renewable Electric Energy”, https://www.fit-portal.go.jp/PublicInfoSummary (2022)

[11] Consumer Affairs Agency https://www.caa.go.jp/policies/council/csic/report/report_012/pdf/report_012_190128_0002.pdf, (2021)

[12] VPN Gate, “Connect to VPN Gate by Using SoftEther VPN (SSL-VPN)”, https://www.vpngate.net/en/howto_softether.aspx (2022)

Firmware Distribution with Erasure Coding for IoT Devices

Takenori Sumi*†, Yukimasa Nagai* , and Hiroshi Mineno†

*Information Technology R&D Center, Mitsubishi Electric Corporation, Japan
†Graduate School of Science and Technology, Shizuoka University, Japan
Sumi.Takenori@dc.MitsubishiElectric.co.jp

Abstract – Due to the increase in the number of IoT devices connected to the internet, 920 MHz frequency bands for wireless communication systems are attracting attention for various IoT applications, e.g., environmental monitoring, smart metering, process monitoring & control. With wireless communication systems on 920 MHz having the features of long distance, low rate and low power consumption, a huge number of IoT devices distributed in wide area can be connected to communication networks. When distributing the same data to IoT devices such as firmware distribution during operation, improving the efficiency of distribution method becomes an issue. We propose a new firmware distribution method with erasure coding for IoT devices. Our computer simulation result shows that proposed method improves the efficiency of distribution by 1.7 times compared with conventional method and achieves higher spectrum efficiency.

Keywords: IoT, Firmware Distribution, Erasure Code

1 INTRODUCTION

In addition to smart phone, laptop and tablet connecting to the internet, IoT devices such as sensors are getting connect to the internet. Due to the increase in the number of IoT devices connected to the internet, 920 MHz frequency bands for wireless communication systems are attracting attention for various IoT applications, e.g., environmental monitoring application for location, temperature, humidity and water lever, smart metering application, process monitoring & control application. Wireless communication systems on 920 MHz (Sub-1 GHz) have the features of long distance, low rate (several 10 kbps – 100 kbps) and low power consumption for conventional standards, e.g., IEEE 802.15.4g, LoRa WAN, SigFox, but the higher data rate up to several Mbps is also considered for applications such as infrastructure monitoring, surveillance camera in IEEE 802.11ah / Wi-Fi HaLow. For long life IoT devices, the firmware update for a large number of IoT devices is also considered as new IoT application. Since 920 MHz has the features of long distance from several 100 m to several km, a large number of IoT devices are deployed in the area. Thus, firmware distribution for IoT devices using 920 MHz with narrow band is a challenge. Thus, efficient firmware distribution method for a large number of IoT devices should be considered. In this paper, we focus on 920MHz for IoT applications and propose the new firmware distribution method using erasure coding to achieve higher efficiency for limited radio frequency.

The rest of this paper is organized as follows. Section II presents related work. Section III describes the proposed firmware distribution with erasure coding. Section IV shows the simulation architecture and results for various conditions. Finally, we conclude our paper in Section V.

2 RELATED WORK

There are existing researches for wireless communications using 920 MHz and firmware distribution. Since 920 MHz is narrow bands compared to 2.4 GHz and 5 GHz for ISM band, special regulation for “10 % transmission duty cycle” and “longer backoff mechanisms” are applied in Japan.

Throughput performance has been demonstrated in [1] and [2], which focus on the PHY and MAC protocol enhancement for higher-throughput, protocol efficiency and delay via simulation and measurement result using prototypes. For example, D. Hotta et al. introduce the performance of multi-hop routing construction using Wi-SUN FAN (Field Area Network) prototypes based on IEEE 802.15.4g FSK PHY [3].

Japanese standard ARIB STA-T108 (20 mW, unlicensed) defines the use of IEEE 802.15.4g system from 920.5 – 928.1MHz (7.6 MHz bandwidth), but the ARIB STA-T107 (250 mW, passive system) and the ARIB STD-T108 (250mW, licensed/registered) also define operation from 920.5 – 923.5 MHz (3.0 MH). Therefore, 923.5 – 928.1 MHz (4.6 MHz bandwidth) is the only reasonable frequency band for IEEE 802.15.4g applications in the unlicensed spectrum. IEEE 802.15.4g is regulated to operate over 200 kHz bandwidth channel in the Sub-1 GHz band. Even low duty cycle constraint applied in the Sub-1 GHz band, e.g., Japanese and European standard allow up to 10% transmission duty cycle [4] - [7] for the number of IoT Devices increased with various standards. Therefore, ensuring higher efficiency for spectrum use in the Sub-1 GHz is clearly important.

3 PROPOSED METHOD

3.1 Challenges in Firmware Distribution

Firmware distribution to IoT devices requires sending the same firmware data to a large number of IoT devices via wireless communication. Firmware distribution using unicast is inefficient because the same firmware data is sent to each device individually. Therefore, it is considered more efficient to distribute to a large number of IoT devices using broadcast. However, it is difficult to deliver all firmware data to all IoT devices in a single communication because wireless communication suffers from packet errors due to low received signal power and/or interference from other wireless systems. As shown in Figure 1, packet errors occur randomly in IoT devices, requiring large number of packets to be retransmitted. Therefore, even when broadcasts are used for firmware distribution, there are issues with distribution

efficiency. In this paper, we propose a method to reduce the number of transmitted packets in firmware distribution by using erasure coding.

3.2 Erasure Code

In this paper, RC QC-LDPC (Rate Compatible Quasi-Cyclic-LDPC)[8] is used as Erasure Code for firmware distribution. As shown in Figure 2, RC QC-LDPC generates redundant packets by erasure coding information packets that divide the data to be transmitted into packets of a certain data length. The data length of information packets and redundant packets are the same. The number of information packets, K, is a multiple of 36, and the number of redundant packets, M, is equal to K or twice as large. RC QC-LDPC can attempt to decode the transmitted data if the total number of received information packets and redundant packets is K or more. Even if some of the received information packets are missing, the transmitted data can be recovered using the redundant packets. Figure 3 shows the relationship between "redundancy rate" and decoding success rate. Here, the "redundancy rate" is defined as the sum of the number of information packets K_{rx} and the number of redundant packets M_{rx} at decoding and the original number of information packets K, using the following formula.

$$Redundancy\ Rate = \frac{K_{rx}+M_{rx}}{K} \quad (1)$$

Figure 3 shows that RC QC-LDPC does not ensure successful decoding even when the redundancy is greater than 1. The higher the number of information packets K, the higher the decoding success rate for the redundancy rate. However, the larger the number of information packets K, the more memory and CPU resources are required for the decoding process.

3.3 Firmware Distribution with Erasure Code

In the proposed firmware distribution method, as shown in Figure 4, the transmitter first erasure codes the firmware data (K information packets in the figure) to generate M redundancy packets (① in the figure). Then, the Transmitter sends up to a total of K information packets and redundant packets (② in the figure). After K packets have been sent, depending on the number of packet errors in the IoT device, additional information packets and redundant packets are sent until the IoT device can decode the firmware data (③ in the figure). Here, the additional packets to be sent are those not sent in ② in the figure. The IoT device uses the received K or more packets to decode and recover the firmware data (④ in the figure). As described in section 3.2, RC QC-LDPC may fail in decoding. If decoding fails, additional packets are sent from the transmitter and the IoT device performs the decoding process again.

To implement the proposed method, IoT devices need to inform the transmitter which packets are missing. In the proposed scheme, IoT devices notify the transmitter of a NACK (Negative Acknowledgement) packet and an ACK

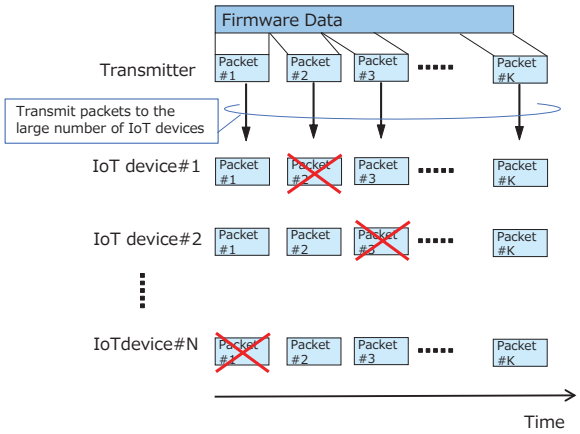


Figure 1: Firmware Distribution for IoT Devices

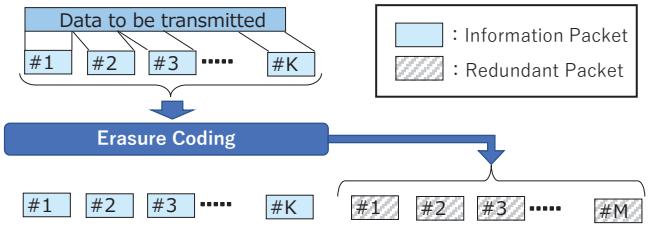


Figure 2: Erasure Coding

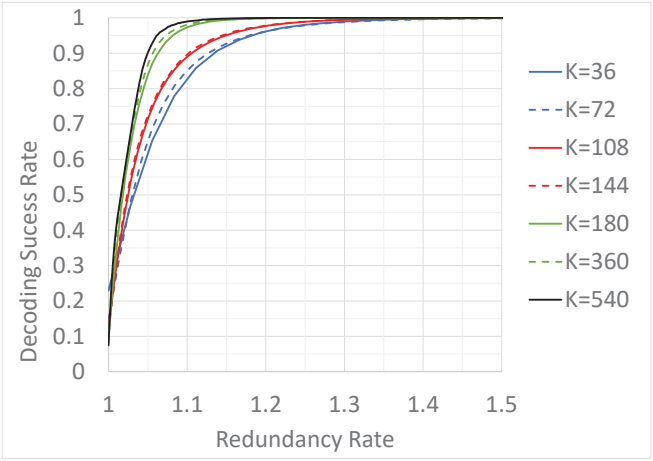


Figure 3: Redundancy Rate vs. Decoding Success Rate

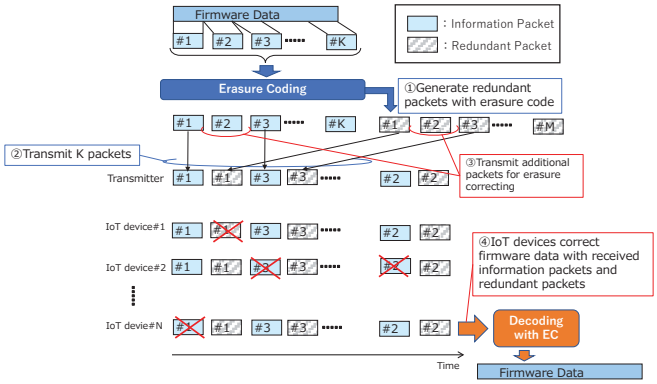


Figure 4: Proposed firmware distribution with EC

(Acknowledgement) packet in the sequence shown in Figure 5. The NACK packet is used to notify which packets are missing; the ACK packet is used to notify that decoding was successful. The packet formats of NACK and ACK packets are shown in Figure 6. UDP is used to send NACK and ACK packets, and Figure 6 shows only the UDP payload. The NACK packet contains a packet type indicating whether it is NACK or ACK and a bitmap indicating packet loss. The length of the bitmap is the sum of the number of information packets K and the number of redundant packets M. The ACK packet contains only the packet type.

4 SIMULATION EVALUATION

We evaluate the proposed method using computer simulation. Table 1 shows simulation parameters. In the computer simulation, IEEE 802.15.4g [9], which is used in IoT devices in the 920 MHz band, was used as the wireless communication method and evaluated in an environment where IoT devices are connected in a star network from a transmitter. The firmware data is divided into 248 Bytes each and broadcast to the IoT device as the payload of UDP packets. RC QC-LDPC was used as erasure code, and the number of information packets K and redundancy packets M were set to 36 - 540. In the computer simulation, the conventional firmware distribution without erasure code also broadcasts the firmware data to the IoT devices by dividing it into 248 Bytes each, as in the proposed method. Then, after the K-packet transmission is finished, the IoT devices send NACK or ACK to the transmitter to notify whether it needs to retransmit data. This is to ensure that the opportunities to send NACK and ACK packets are the same for the conventional and proposed methods. In the conventional method, the number of redundancy packets M is 0 because erasure coding is not performed for the firmware data.

First, the relationship between the number of information packets K and effective throughput is shown in Figure 7, which shows the simulation results when firmware is distributed to 20 IoT devices and PER is 10%. Since a PER of 10% or less is often set for wireless communication systems to take operations into account [9], the simulations in this paper are based on an evaluation at a PER of 10%. According to the figure, the throughput of the proposed method is higher than that of the conventional method, regardless of the value of K. As the value of K increases, the effective throughput of the proposed method increases. As mentioned in section 3.2, this is because the larger K is, the higher decoding success rate at the same redundancy rate. However, since the memory and decoding processing load increases as K increases, the subsequent simulation evaluation will be performed for the case where the number of information packets K is 180, where the effective throughput increase rate by the proposed method is relatively high.

Next, Figure 8 shows the simulation result on the relationship between PER and effective throughput when firmware is distributed to 20 IoT devices with the number of information packets K set to 180. The figure shows that if the PER is 1% or higher, the effective throughput is higher with the proposed method than with the conventional method. However, the effective

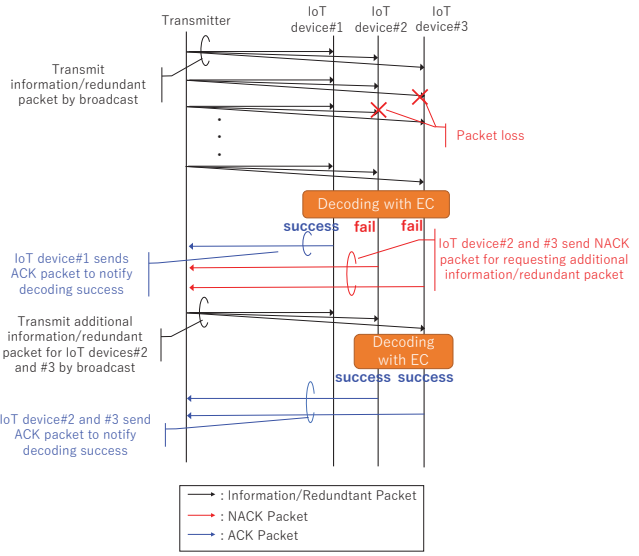


Figure 5: Packet sequence of the proposed method

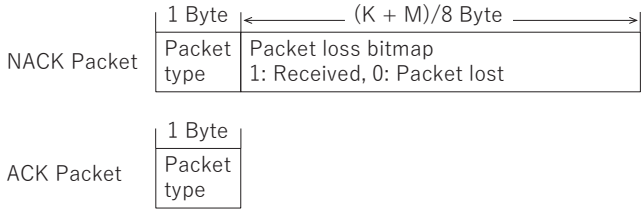


Figure 6: Packet formats of NACK and ACK packet

throughput at a PER of 0% with no packet errors is approximately 40 kbps for the conventional method, but 33 kbps for the proposed method, which is lower than the conventional method. As mentioned in Section 3.2, this is because RC QC-LDPC codes may not succeed in decoding even if a total of K or more information packets and redundancy packets are received. Additional packets need to be sent from the transmitter to the IoT devices. Since packet errors occur in wireless communication systems, the proposed method can be applied to improve the effective throughput in an environment with a PER < 10%, where wireless communication systems are normally operated.

Finally, the relationship between the number of IoT devices and effective throughput is shown in Figure 9, which shows simulation results when the number of information packets K is 180 and PER is 10%. The figure shows that the effective throughput of the proposed method is higher than that of the conventional method when firmware distribution is performed to two or more IoT devices. In the case of a single IoT device, the effective throughput is lower than the conventional method due to the significant impact of the possibility of unsuccessful decoding even if the number of received packets is K or more. For 20 IoT devices, it is 1.60 times the effective throughput of the conventional method; for 50 devices, it is 1.73 times; and for 100 devices, it is 1.76 times. However, as the number of IoT devices increases, the effective throughput decreases: 29.7 kbps for 20 IoT devices, 24.8 kbps for 50 devices, and 19.8 kbps for 100 devices. This is likely due to an increase in NACK packets.

The transmission duty cycle of the transmitter in the simulation of Figure 9 is shown in Figure 10. As mentioned

in Chapter 2, Japanese and European standard allow up to 10% transmission duty cycle. Figure 10 shows that regardless of the number of IoT devices, the transmission duty cycle of the transmitter exceeds 10% for both the proposed and conventional methods. The transmission duty cycle of the proposed method is lower than that of the conventional method, and it tends to be lower when the number of IoT devices is larger. Therefore, in Japan and Europe, when the transmission duty cycle is 10% or less, the difference between the effective throughput of the proposed method and the conventional method is expected to widen.

5 CONCLUSION

We proposed the new firmware distribution method using erasure coding to achieve higher efficiency for limited radio frequency. The performance of the proposed method was evaluated by computer simulation and compared to the effective throughput of conventional firmware distribution methods. By applying the proposed method to firmware distribution, the effective throughput was found to be 1.60 times higher than that with the conventional method when there are 20 IoT devices receiving firmware data, 1.73 times higher when there are 50 devices, and 1.76 times higher when there are 100 devices. In this paper, we evaluated the performance of the proposed method in a single-hop wireless network. Our future work is to investigate the application of the proposed method to multi-hop wireless networks.

Table 1: Simulation parameters

Firmware distribution parameters	Value
The number of IoT devices	1 - 100
Erasure code	Rate-Compatible QC-LDPC
K, the number of information packets	36, 72, 108, 144, 180, 360, 540
M, the number of redundant packets	Conventional method: 0 Proposed method: same as K
Firmware distribution packet length	248 Byte
Firmware distribution ACK packet	1 Byte
Firmware distribution NACK packet	1+(K+M)/8 Byte
6LowPAN and UDP header	53 Byte
PHY/MAC	IEEE 802.15.4g
MAC parameters	Value
macMinBE	5
macMaxBE	8
LIFS	1000 us
aUnitBackoffPeriod	1130 us
phyCcaDuration	130 us
aTurnaroundTime	1000 us
tack	1000 us
MAC header size	11 Byte
FCS	2 Byte
ACK frame size	9 Byte
PHY parameters	Value
Data rate	100 kbps
Modulation	2-FSK
Modulation index	1.0
Frequency	923.7 MHz
Channel spacing	400 kHz

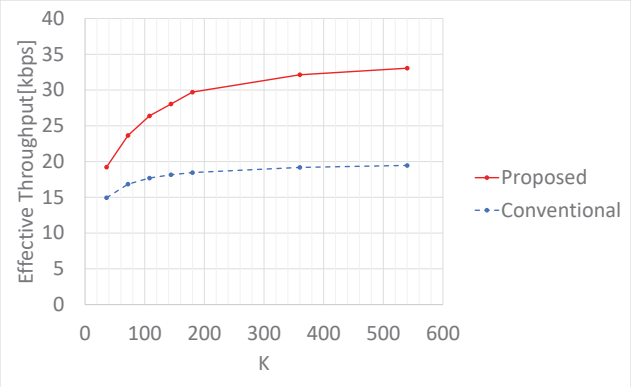


Figure 7: K vs. Throughput (PER: 10%, 20 IoT devices)

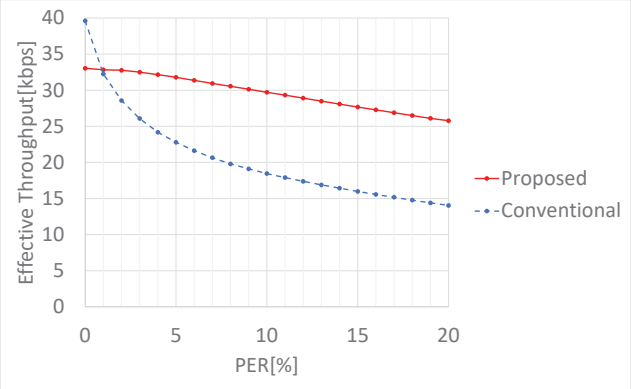


Figure 8: PER vs. Throughput (K: 180, 20 IoT devices)

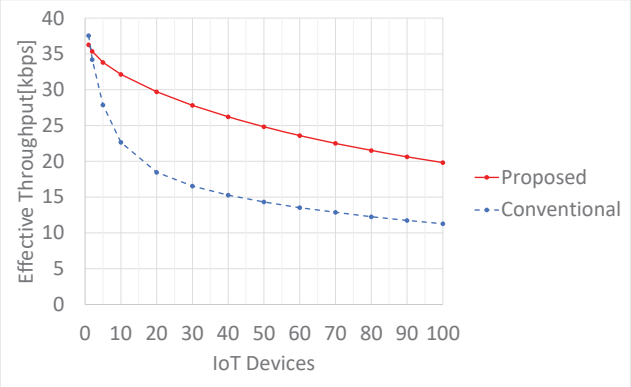


Figure 9: The number of IoT devices vs. Throughput (K: 180, PER: 10%)

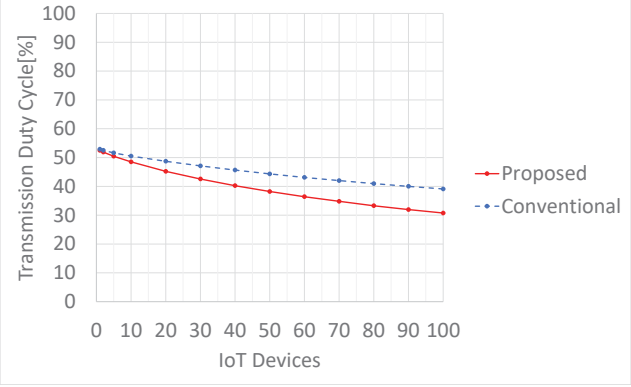


Figure 10: The number of IoT devices vs. Duty Cycle (K: 180, PER: 10%)

REFERENCES

[1] C.-S. Sum, F. Kojima, and H. Harada, ‘‘Performance analysis of a multiPHY coexistence mechanism for IEEE 802.15.4g FSK network,’’ in Proc. IEEE Wireless Communication Network. Conf. (WCNC), Apr. 2013, pp. 41–46.

[2] F. Righetti, C. Vallati, D. Comola, and G. Anastasi, ‘‘Performance measurements of IEEE 802.15.4g wireless networks,’’ in Proc. IEEE 20th Int. Symposium World Wireless, Mobile Multimedia Network (WoWMoM), Jun. 2019, pp. 1–6.

[3] D. Hotta, R. Okumura, K. Mizutani, and H. Harada, ‘‘Stabilization of multi-hop routing construction in Wi-SUN FAN systems,’’ in Proc. IEEE 17th Annual Consumer Communication and Network Conference (CCNC), Jan. 2020, pp. 1–6.

[4] 920MHz-Band RFID Equipment for Specified Low Power Radio Station, ARIB STD-T107 version 1.1, ARIB, Tokyo, Japan, 2017.

[5] 920MHz-Band Telemeter, Telecontrol and Data Transmission Radio Equipment, ARIB STD-T108 version 1.2, ARIB, Tokyo, Japan, 2018.

[6] Y. Nagai, J. Guo, P. Orlik, Ben A. Rolfe, and T. Sumi, Proposal for Section 6.1 Japan on Recommended Practice, document IEEE 802.19-19/0049r0, IEEE 802.19, 2019.

[7] T. Sumi, Y. Nagai, J. Guo, and P. Orlik, 920 MHz Status Update in Japan, document IEEE 802.19-18/0084r0, IEEE 802.19, 2018.

[8] Wataru Matsumoto, Rui Sakai, Hideo Yoshida, “Rate Compatible QC-LDPC codes,” SITA 2006, pp.387-390 (2006).

[9] IEEE 802.15.4-2020, “IEEE Standard for Low‐RateWireless Networks,” IEEE 802.15.4-2020, IEEE, (2020).