International Journal of Informatics Society

06/19 Vol. 11 No.1 ISSN 1883-4566

Informatics Society

Aims and Scope

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

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

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

Guest Editor's Message

Tomoyuki Yashiro

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

We are delighted to have the Thirty-first issue of the International Journal of Informatics Society (IJIS) published. This issue includes selected papers from the Twelfth International Workshop on Informatics (IWIN2018), which was held at Salzburg, Germany, Sept. 9-12, 2018. The workshop was the twelfth event for the Informatics Society, and was intended to bring together researchers and practitioners to share and exchange their experiences, discuss challenges and present original ideas in all aspects of informatics and computer networks. In the workshop 26 papers were presented in seven technical sessions. The workshop was successfully finished with precious experiences provided to the participants. It highlighted the latest research results in the area of informatics and its applications that include networking, mobile ubiquitous systems, data analytics, business systems, education systems, design methodology, intelligent systems, groupware and social systems.

Each paper submitted IWIN2018 was reviewed in terms of technical content, scientific rigor, novelty, originality and quality of presentation by at least two reviewers. Through those reviews 20 papers were selected for publication candidates of IJIS Journal, and they were further reviewed as a Journal paper. We have three categories of IJIS papers, Regular papers, Industrial papers, and Invited papers, each of which were reviewed from the different points of view. This volume includes six papers among those accepted papers, which have been improved through the workshop discussion and the reviewers' comments.

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

Tomoyuki Yashiro received his Ph.D in Engineering from Keio University in 1999. He became a lecturer at Chiba Institute of Technology in 1998 and an associate professor in 2001. Currently, he is a professor of Faculty of Information and Computer Science, Chiba Institute of Technology from 2009. His research interests include the communication system of Intelligent Transportation Systems(ITS) , Location Based Service and so on. He is Fellow of Information Processing Society Japan and members of IEEE and IEICE.

Industrial Paper

Quantitative Risk Management Method using Logistic Regression Analysis

Akihiro Hayashi†**, Nobuhiro Kataoka‡, Yasunobu Kino*, Mikio Aoyama**

†Department of Information Design, Shizuoka Institute of Science and Technology, Japan
‡Interprise Laboratory, Japan
* Graduate School of Business Science, The University of Tsukuba, Japan
** Graduate School of Information Science and Engineering, Nanzan University, Japan
pixysbrain@gmail.com

Abstract -

Recently, the interest in the risk management (RM) process has been growing. RM aims to lead a project to success by eliminating negative factors that can cause it to fail. Therefore, it is expected that the number of failed projects can be reduced in organizations where have introduced RM process. However, this expected result has not been obtained yet. In this study, we first analyze the RM process for the system development projects conducted recently, then we figure out that the issue is that the RM implementation is not in time for actual trigger where RM process introduced successfully. Next, we identify the factors where RM did not meet the expected criteria and propose a quantitative RM method that could improve the RM and project management (PM) process by using Earned Value Management (EVM) and Logistic Regression Analysis (LRA) to eliminate the factors. By applying the proposed method to a real RM case, we concluded that the proposed method is effective.

Keywords: Project Risk Management, Logistic Regression Analysis, Quantitative Project Management

1 INTRODUCTION

We live in a software dependent society in which software plays a major role in various kinds of products such as organizational operation systems, home appliances and automobiles. It means that many companies inevitably focus on system development including a large amount of software.

However, according to reliable statistical information[1], only 27% of projects succeed in all aspects of quality, cost, and delivery time (QCD) in domestic and foreign system development projects. Thus, three-fourths of the projects do not meet all the QCD criteria, which leads to the cancellation of 24% of software development projects[2].

To solve this problem, interest in introducing RM processes in system development has increased. It is believed that RM has the ability to lead projects to success by eliminating negative factors that may cause the project to fail.

To introduce RM processes, an international "best practice model" has adopted the Project Management Body of Knowledge (PMBOK)[3], Program & Project Management for Enterprise Innovation (P2M) [4], the 2nd version of Projects in Controlled Environments (PRINCE2) [5] as reference models, and introduced specific practices for RM that all presented in these guides.

However, even the introduction of the "best practice," does not reduce the number of failed projects. The findings suggest that successful implementation of the RM process does not contribute toward the reduction of failed project occurrence.

We believe that there are two major factors that contribute to the event described above.

One factor is that the project management standards and guides that have been proposed and developed overseas do not match practices that are in place in the domestic system development projects. This is because the standards and guides developed overseas are often based on large-scale projects. For example, the Capability Maturity Model Integration (CMMI) has been developed based on the assumptions drawn from the procurement model of the Department of Defense (DoD) of the United States. The DoD constantly carries out large-scale projects for procurement of munitions. Although it is extremely difficult to apply these assumptions to standard-sized system development projects in Japan, the focus was on the necessity and not on the prevailing practices.

The other factor is that even though the standards and guides are correct, they have not been successfully introduced to the system development site. When standards or guides are introduced to a system development site, conformance to the standards and guides take precedence. However, the goal of these standards and guides is not conformance, but performance.

In both cases, it is necessary to establish an appropriate methodology to introduce RM processes to the management of standard-sized projects in Japanese industry.

To address this issue and decrease the incidence of failed projects, we first analyze four cases of a specific risk management process conducted recently and identify the factors that will create a bottleneck. Next, to solve the problem of failure, we proposed a method to introduce the RM process appropriately. The proposed method included quantitative risk management and the implementation of risk countermeasures. When we applied our proposed method to a real case for the RM of system development projects, a measurable effect was observed in the form of a reduction in the number of failed projects and a reduction in the contingency budget.

The remainder of this paper is organized as follows: In Section 2, we review the related works on PM, EVM and RM to confirm the originality of this study. In Section 3, we analyze a case on the implemented RM practice where we consult and identify the reason behind the failure of RM, and describe the issues to be solved. In Section 4, we propose a quantita-

tive RM method by using a statistical tool LRA. In Section 5, the effectiveness of the proposal is evaluated by applying the method to a real case of system development. In Section 6, we discuss the results of the case study. In Section 7 presents the conclusion.

2 RELATED WORKS

The first step in designing a research strategy involves specifying the research question. The research question is how to establish a new methodology for RM that involves the application of quantitative PM and EVM for performing risk corrective/preventive actions. The second step is to apply this methodology to the real case on RM and evaluate the effectiveness of the proposed method.

Hereafter, we review the prior research in the field of RM, PM, and EVM.

2.1 RM

Project risk is defined as an uncertain event or condition that, if it occurs, would have a negative effect on one or more project objectives such as scope, schedule, cost, or quality [3].

In most of the earlier research on RM of system development, Boehm[6] and Williams[7] initially introduced the implementation methods of RM practices that are commonly practiced, such as risk identification, evaluation, classification, and prioritization. In this methodology, after the establishment of the basic RM basic technique, the researchers attempted to optimize the project schedule by considering the simultaneous effect of the risk associated with one task on the other risks factors; these factors are proposed in a quantitative framework of analysis for supporting decision making in project risk response planning.

They used a design structure matrix representation to capture risk interactions and build a risk propagation model for predicting the global mitigation effects of risk response actions. Unlike the technique proposed by the author, they did not exploit the potential of the methodology for exploring the impact of the risks on a single activity, thereby neglecting how the network topology could change in relation to the risk propagation and determine possible project delays.

Acebes et al. [13] proposed a methodology to integrate EVM with risk management, based on traditional EVM indicators that allow project managers to detect negative deviations from planned values, corresponding to cumulative positive or negative cost/schedule buffers. Such information can be usefully employed to take corrective actions or identify the sources of improvement and further optimize project activities.

Muriana et al. [9] explained a deterministic technique for assessing and preventing project risks, by determining the risk of the work progress status. As each phase ends, the actual value of the input factors are detected and compared with those of the planned values, and corrective actions are taken for considering the impact of the actual performances on the overall project. The current risk degree of the project is determined through the weighted sum method. If it is higher than planned, then preventive actions are taken to mitigate the risk of the entire project. However, the authors limited their work to the determination of the deviations from planned values, without focusing on preventive/corrective actions that can be put into practice.

2.2 PM

The origin of modern PM can be dated between 1900 and the 1950s [8]. Before the 1950s, the focus of PM techniques was primarily on scheduling, that is, the understanding of activities and sequencing. PM is a critical activity that determines the success or failure of a project. Therefore, several techniques have been perfected over time to simplify the efforts related to such an activity and increase its usefulness.

The first attempt to support project managers in the scheduling phase was made with the introduction of the Critical Path Method (CPM) and Program Evaluation and Review Technique (PERT); additionally, these methods were invented and introduced in the 1950s. However, CPM posed uncertainties regarding which deterministic techniques allow the determination of the longest path in the network named “critical path” and which tenure is taken as the earliest time for project completion. The PERT was introduced some years later, adding to the hypothesis of uncertainties regarding the activity duration.

After the dawn of this era, a lot of new approaches of PM were proposed and attempted. Some works have focused on communication and coordination issues in projects. Curtis et al. [14] studied how communication networks and breakdowns affected a development project. Their findings raised many issues that are critical for awareness systems, such as the importance of both formal and informal communication in development. Gutwin et al. [15] looked at several open source development projects and found that developers needed to have an awareness of other factors to contribute toward development and that the developers gained an awareness primarily through text based communication. Herbsleb and Grinter [16] conducted a field study showing the importance of informal communication; furthermore, the difficulty in communicating across globally distributed teams suggested that an increase in awareness would benefit development.

Others have looked at the mismatch between coordination requirements and actual communication [17], [18] and proposed mechanisms to improve the mismatches that occur. Espinosa et al. [19] have identified factors that affect awareness in software development, including awareness about the nature of team knowledge and distance.

Finally, a simulation method has proposed a technique to manage project scheduling [10], rapidly becoming one of the most-used techniques for large-sized projects affected by uncertainties emerging from the activities.

2.3 EVM

In 1962, the Work Breakdown Structure (WBS) was invented. In particular, it provides project managers with techniques for monitoring a project through the employment of EVM. EVM was introduced in 2000 in the PMBOK guide and is today broadly employed in the field of PM for measuring project performances [11] because it combines measure-

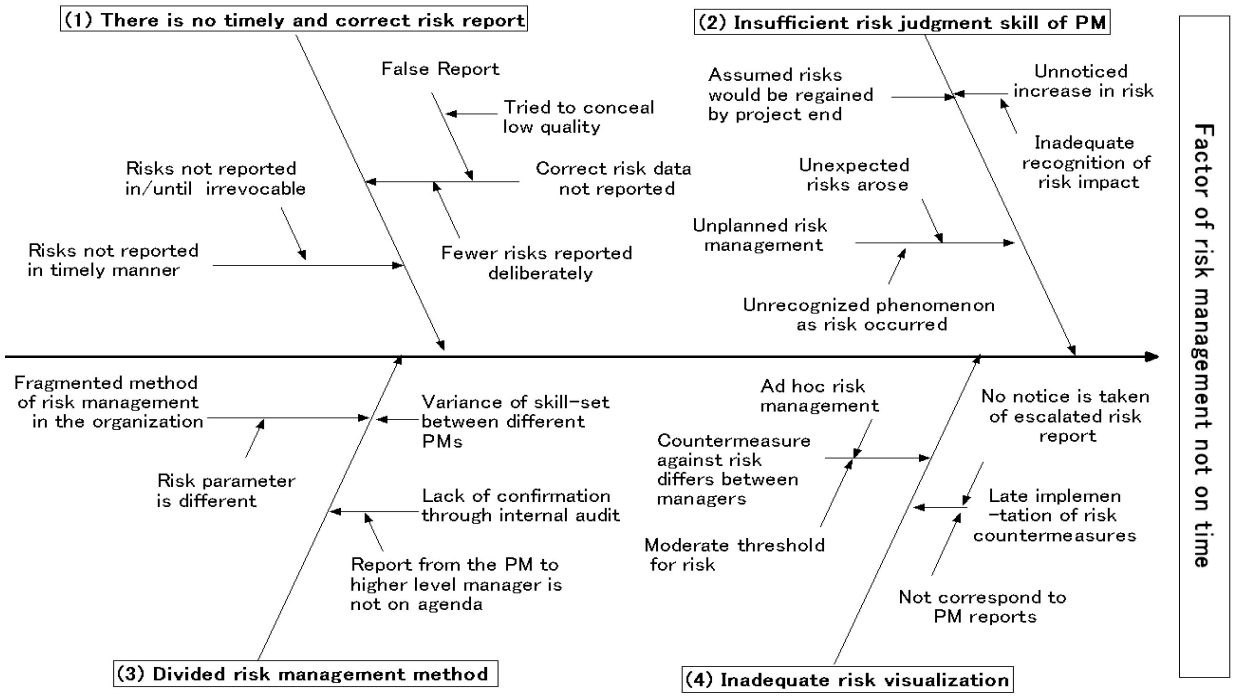


Figure 1: Fish-Bone Chart of Problems.

ments of the Iron Triangle of the PM. The usefulness of EVM in forecasting the project performances is widely recognized, and considerable research has been published that attempts to extend this technique.

The introduction of EVM technique was followed by studies that proposed a method similar to that of quantitative management. For example, Pajares et al.[13] proposed two new metrics that combine the EVM and Project RM for project controlling and monitoring. The study compares the EVM cost and schedule variances with the deviation that the project should have under the expected risk analysis conditions that allow project managers to analyze whether the project overruns are within the expected variability or there are structural and systemic changes over the project life cycle.

Deshpande at el. [21] compared correlation and regression coefficient using three distributions. Function exaction using correct distribution for forecasting project duration and cost will prevent significant losses in future. Therefore, an attempt is made to find the alternative distribution of cost performance index (CPI) and schedule performance index (SPI) for better decision making. If the project schedule performance shows poor results, then it would be essential for a manager to take corrective actions with the help of this tool.

Although such prior research explains the potential of success for the RM, it does not describe how to establish a methodology that can be applied to the system development project in the real world. To the best of our knowledge, no particular prior research discusses the appropriate methodology for introducing the risk management process.

3 FACTORS CONTRIBUTING TO THE FAILURE OF RM

3.1 Case Analysis of RM Process

First, we aim to clarify the reasons due to which the number of failed projects has not decreased, even after best practice of the successful introduction of the RM process in organizations. We analyze the factors by taking up four organizations for whom we have provided management consultation to date. Below is a summary of the organizations to be analyzed:

Case 1 Electronic control of vehicle amenity
Case 2 Electronic notebook, which maintains a schedule, dictionary, calculator, and custom program
Case 3 Air conditioner system controlled by an internet-based remote control
Case 4 Derivative development of value-added of acoustic measurement calibration equipment

By analyzing the 4 aforementioned development cases, we found the following four problems in the RM process:

Problem 1 Since the triggering of alarm for the notification of risk was delayed by the project manager (PM), risk countermeasures could not be implemented on time. A similar problem occurred multiple times in a PM's tenure; he thought that the problem could be solved every time. Therefore, he did not report the emergence of risk to the higher-level managers.
Problem 2 Since the development project was originally planned the development period, it was biased toward keeping the delivery date. Therefore, the organizations

Project Management PDCA

	Plan	Do	Check	Act
Project Manager	Making Risk Management Ledger and WBS		Risk Analysis using LRA	Activate Risk Counter Measure
Development		WBS update weekly RM Ledger update process breaks point		
Risk Management	PV=EV	As always EAC/BAC > 100 %	If Recoverable % > (EAC/BAC -100%)	Drive Recoverable % < (EAC/BAC -100%)

Figure 2: Procedure of this proposal.

were averse toward reworking due to RM activities and hence hesitated to report risk occurrence.

Problem 3 Despite the original plan to activate risk response measures in an event-driven manner, project members did not accurately understand the RM process and subsequently reported risks at weekly progress meetings that caused notification delays.

Problem 4 Despite a clear definition of the trigger and threshold for risk interpretation in the RM ledger, the roles and responsibilities were misunderstood, and the risks were not reported correctly.

The four problems listed above indicated that the RM practice was correct, but that risk countermeasure actions were not implemented on time.

3.2 Factor Analysis to show that RM was not Implemented on Time

Next, we analyze factors that contributed toward the delayed occurrence of RM by creating a fish-bone diagram, as shown in Figure 1, extracted from the documents and minutes of the meeting of the past RM assessment. As a result, the following four factors were clarified:

Factor 1 There was no timely and correct presentation of a risk report:
Even when the risks expanded and severe delays did not allow adequate management, organizations sometimes reported less risk or kept critical risks hidden since the project was at a stage wherein it would be evaluated by higher-level managers. Therefore, a correct risk report was not delivered on time. This is the cause of Problem 1.

Factor 2 Insufficient risk judgment skill of the PM:
Due to the PM's insufficient risk judgment skill, such as insufficient identification of risk and undistinguished critical risk, the organization failed to manage the risk properly. This led to Problems 1, 3, and 4.

Factor 3 Divided RM method:
Due to insufficient communication between the PM and higher-level managers, unclear terms and methods were used, and insufficient information was presented, which contributed to Problems 2 and 3.

Factor 4 Inadequate risk visualization:
The RM ledger included the PM's subjective evaluations. Higher-level managers were unable to monitor risk situations of the projects, which contributed to the Problem 4.

4 RM METHOD USING QUANTITATIVE PROCESS MANAGEMENT

4.1 Basic policy

A risk is a potentiality thing, and it does not necessarily become explicit. Therefore, the activation of risk countermeasures ahead of schedule will lead to the employment of unnecessary labor and costs.

For establishing a good RM process, it is important to define the risk of each project and judge them objectively by using quantitative data.

(1) Set a clear risk criterion
Objective judgment criteria are set for practicing each RM process, such as registering in the RM ledger, identifying risk explicitly, and using the quantification method.

(2) Process Performance Baseline (PPB)
With an emphasis on historical project data, not focusing on each project database, but the PPB focus on all historical projects' data accumulated.

(3) Evaluating the entire project status using a statistical method
The individual risk threshold is not evaluated, but whole project threshold is evaluated using a statistical method.

(4) Introduction of subject matter expert (SME) and quality assurance (QA): To solve the problem of skill shortage related to RM, an SME, a specialist of the development

No	Risk Description	Probability	Impact	Rework	Response	Threshold	Countermeasure	Priority
1	Delivery from outsourcing companies tends to be delayed	H	H	100	Mitigate	May 11th	add resource	3
2	The quality of some project member 's program is always bad	M	H	100	Mitigate	Design Process	assign mentor	4
3	Influenza will become prevalent and absentees will appear	L	L	300	Mitigate	RA Days	Reserve party	1
4	The president may change and the policy may change	L	M	200	Mitigate	Gen meeting of shareholders	New Strategy	2
	H:High, M:Middle, L:Low		Total	700				

Figure 3: RM Ledger.

process who belongs to the engineering field, and a QA are introduced to discuss risks activities.

(5) Alignment Aligning the basic policies of (1) to (4) with the RM process.

4.2 Procedure

In order to solve the issue concerning a delay in the RM, which was identified in section 3.2, we will introduce the following RM procedures: making RM ledger and WBS, quantitative progress management using EVM, risk analysis using LRA, and risk counter strategy. These RM Procedures are shown in Figure 2. Hereafter, we will explain the RM procedures in detail.

4.2.1 Making RM Ledger and WBS

In Stage 1, we first make RM ledger and WBS. In the system development project, risks of the entire project are identified at the project planning stage. Mainly, risks are obtained as a result of awareness created during the process of analyzing customer need, creating customer requirement definition, creating project planning, and reviewing the QA Document.

The obtained risks are listed in the RM ledger of the project, and the properties of each risk are set as shown in Figure 3. These properties include risk description, probability, impact, rework time, risk response strategy, the threshold of action taken, risk countermeasures, and priority. Priority is that the magnitude of the risk influence is sorted by order.

At the project planning stage, we also make WBS. WBS is a key project deliverable that organizes the team's work into manageable sections by hierarchical decomposition of the work to be executed by the project. At the project planning stage, we review all the tasks and set start and end dates of each task and efforts that are to be spent on the task.

When we identify all the tasks and make the WBS of a project, we automatically know the EVM value of the current status because the planned value (PV) and budget at completion (BAC) are calculated entirely.

In the RM process, the influence of risk is converted to "time" or "money." In this study, we convert the influence to "time" (minutes). The project stakeholders can understand the amount of influence quantitatively.

4.2.2 Quantitative Progress Management using EVM

In Stage 2, the project is managed on a weekly basis. Usually, a progress management meeting is conducted on a weekly basis. The project manager asks project members in charge to update WBS for the concerned week at the meeting. Subsequently, the actual value compared to the estimated value achieved in the project is known for a particular week. At the meeting, project members report the manager's current progress status by calculating these EVM values and problems that occurred, if any. This enables the managers to estimate the project's total cost at project completion, which is also referred to as estimate at completion (EAC)

In system development, by using a waterfall model, it has been empirically found that the risk is often explicit at process breaks. Therefore, at the progress meeting that is held at the end of the process, project members and SME conduct a risk review meeting. They reevaluate the risks according to the change of the environment at that time and update the RM ledger.

4.2.3 Risk analysis using LRA

At stage 3, we can predict whether the project will fail in the future, by using LRA. LRA is a statistical method that predicts the occurrence probability of an event from the size of accumulated data.

When we use a risk value as an explanatory variable, value that can take only a binary response (Yes / No), like the occurrence of project failure as a dependent variable, the probability of the influence on the occurrence of the failed project can be determined.

According to the basic policy (3), we decided that the whole project risk, instead of individual risk, should be set as the progress management threshold. Therefore, we decided that EAC/BAC should be set as the criteria for evaluating a project's success or failure. When the EAC/BAC exceed the specific trigger, the project manager should take appropriate action under the RM.

Organizations that conduct system development projects often have similar degrees of difficulty and similar scales. We have created repositories of PPB by accumulating project data over the past several years. We can also quantify the recoverable period for each construction period if the project is delayed.

For example, it is empirically known, "If you are projecting for 18 months, you will recover and meet the delivery date if the progress delay is less than 5% of the entire project period." We could calculate the recoverable period for each project by employing the LRA. It was found that if the period exceeds the value of (EAC/BAC- 100%), then it will lead to a delay in the delivery date. Subsequently, it will be necessary to take countermeasures that will not make the risk manifest in the schedule of recoverable limits.

4.2.4 Risk Counter Strategy

In stage 4, the project manager monitors the risk status at a progress meeting and takes appropriate risk actions if necessary. If the result of LRA exceeds the threshold, then the project manager can compensate for the project delay by immediately activating risk counter measures according to the priority measures set in the RM ledger until the EAC/BAC comes below the value EAC/BAC-100%. Then let the next PDCA cycle start.

5 EVALUATION OF THE APPLICATION IN ACTUAL DEVELOPMENT ORGANIZATION

In Section 5, a case study wherein the proposed RM method is applied at Company A and its effectiveness is evaluated. For the application of LRA, JMP®14 (SAS Institute Inc., Cary, NC, USA), which operates on Windows PC, was used.

5.1 Case Study for Embedded System Development

Company A introduced project management using PMBOK for about 10 years. Development projects comprising the basic operation of the project management using the PDCA cycle, is well established.

However, in reality, even though it was called the RM process, its focus was on creating a risk matrix and completion record for the preparation of assessment evidence. This is a situation that does not lead to effective RM.

Company A set the development process standard, based on the waterfall model that considered the entire organization. Additionally, at the time of the introduction of the PMBOK, the RM plan, the creation of the RM ledger, and the risk assessment checklist were managed within the company. The company mainly undertook derived development. The company focuses on delivering in a timely manner in a short cycle. In the cases where the delivery is delayed among QCDs, the project is labeled as a "failed project." The policy of the organization is to prevent delayed delivery.

Company A's customers do not present their requirements clearly. Therefore, the project manager in Company A must analyze customer needs and make a feasible completion plan, and formulate the budget and set the construction duration accordingly.

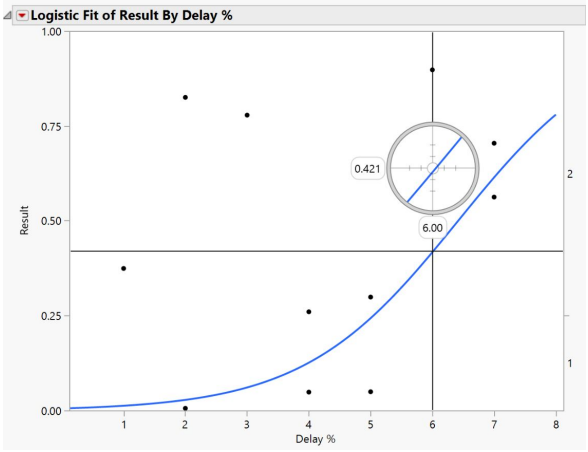


Figure 4: The Causes of Failure and its Proportion.

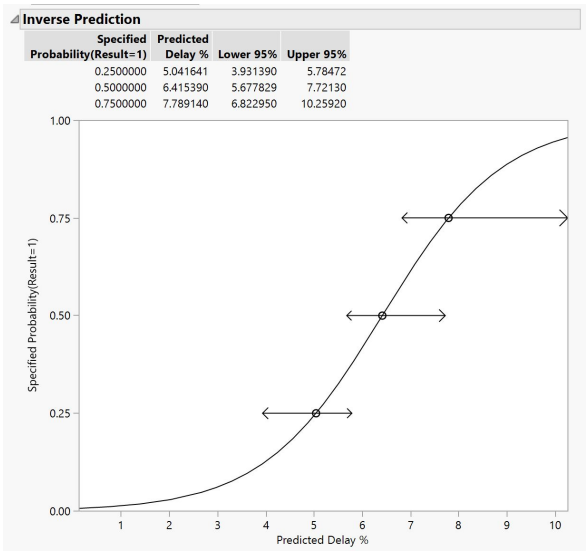


Figure 5: LRA and Inverse Estimation.

Subsequently, to implement the requirement definition, the company undertakes project planning and conducts a QA review for implementing the requirement definition and project plan.

After a discussion with the PM, SME, QA, and project members, the company rules out the related risks and prepares the RM ledger based on the risks.

The project manager holds a progress meeting on a weekly basis. The project manager asks all the project members in charge to update the WBS. Subsequently, the project manager calculate the projected EAC and EAC/BAC. Then they update the RM ledger with latest data and entire EVM value is converted into "time."

The development period at Company A is set at around 6-18 months, depending on the scale of development. We calculate the probability of project failure by using the LRA.

For example, when a project construction term is 18 months, the predicted probability of failure would be as shown in Fig-

ure 4. The cross-hair tool indicates that the prediction probability is 0.421 at the time of 6% delay. In other words, if the project is delayed by 6%, then the delivery date will be delayed with a probability of 42%.

A 42% probability is difficult to employ as a psychological milestone to activate risk countermeasures. A value at which the prediction probability becomes 50% was calculated using an inverse estimation of LRA, and it was 6.41%. Conversely, it means that "If the project is delayed by 6.41%, then there can be a 50% chance of a recovery." This state is shown in Figure 5.

Actually, if the recoverable range exceeded the threshold (like 6.41%), they activated risk countermeasures to reduce the project delay in descending order of influence until the value fell below the threshold to ensure that the value fit within the recoverable range.

5.2 Evaluation of Effectiveness

Figure 6 shows the three-year trend of delayed project delivery in Company A. Although the duration for which the project delays were observed is small, the number of projects subjected to delivery time delay has definitely been reduced. Meanwhile, Company A did not introduce any other measures during this time, but the proposed method has been introduced. Company A considered the transition, shown in Figure 6 which this can be regarded as an improvement that is achieved through the proposed method.

Figure 7 shows the transition of contingency for 3 years, after the introduction of the proposed method in Company A. Contingency is a reserve expenditure fund that can be drawn on to prevent project settlement deficit. It is preferable not to use contingency funds because it is recorded as a profit if not used. In the first year, after the introduction of our proposed method, we consumed nearly 30% of the contingency. It was suppressed to 20% or less in the third year. Since countermeasures are given priority in the order of the risk of damage due to anticipated risk, we believe that it contributed to the prevention of major deficit in projects. We confirmed that the proposed method also improves cost.

6 DISCUSSION

In this section, we discuss whether the factors presented in section 3.2 have been resolved.

6.1 Timely and Correct Risk Report

In this study, project risks at each process break are identified at the project planning stage and reviewed by PM, SME, and QA during the RM meeting. The RM ledger and WBS are updated. Subsequently, we obtain the EVM value and objectively calculate appropriate parameters for the project. At these meetings, the overall project risk and progress status are reported in a timely manner. The project manager can judge the status promptly.

Thus, nobody can reduce the number of risks or hide critical risks while reporting them; the risks are reported in an accurate and timely manner, thereby resolving the issue of delayed and erroneous risk reports.

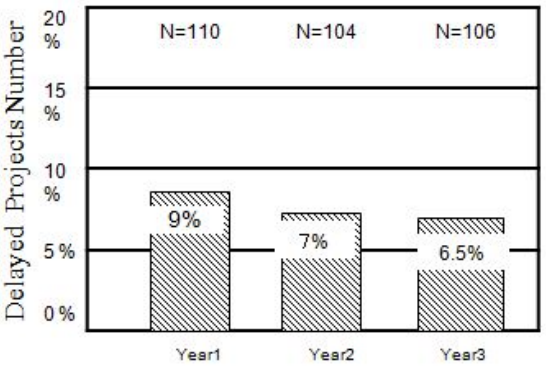


Figure 6: No. of Delivery Delayed Projects.

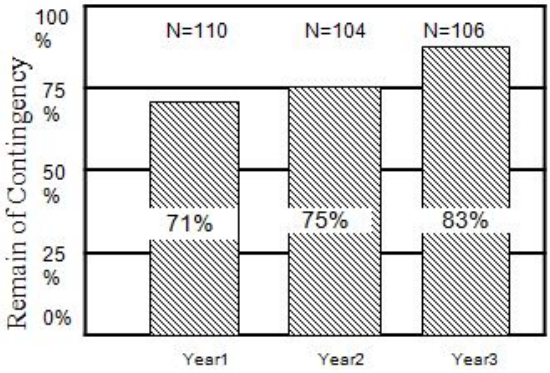


Figure 7: Transition of Contingency Budget.

6.2 PM's Risk Judgment Skill

In this study, to identify and judge risks, an SME and a QA reviewer are assigned to the project. Instead of a project manager, they support and perform RM processes, including risk identification and the activation of risk countermeasure.

In addition, risk is evaluated objectively by using parametric data to avoid the biases of project managers that might result from their assumed expertise over all the technical fields. It implies the resolution of the issues concerning skill shortage in project managers and their lack of risk-judgement skill.

6.3 RM Method in an Organization

In this study, the organization's historical project data are accumulated through the PPB. Recoverable range of each project period calculated by LRA method are accumulated.

Project managers or higher-level managers can judge risks objectively and activate risk countermeasures. The proposed RM method facilitated the unification of the organization's RM method. Thus, this method contributed toward the effective implementation of the RM method in an organization.

6.4 Full Visualization of Risk

In this study, after identifying the risk through an upstream process, we monitored the risks through weekly meetings and visualized the magnitude of an allowable recovery range by using statistical methods. Furthermore, the magnitude of the

risk effect was converted into "time." We visualized the possibility of leadtime to delivery delay. Thus, the issue of full visualization of risk was resolved.

7 CONCLUSIONS

In this research study, we analyzed four actual RM processes carried out at the system development site. Additionally, the study identified the factors that contributed to delayed RM, despite the introduction of the correct RM process at the organization. Subsequently, we proposed the RM method using a quantitative process management approach that included PM, EVM, and LRA.

When we applied the method to the actual embedded development projects, we could verify and confirm the improvement effect on the reduction of the number of projects with delayed delivery times and a decrease in contingency. Thus, the proposed method is considered potentially effective.

This proposed method can be introduced easily in any organization implementing process improvement. In the future, it is necessary to increase case examples, evaluate effectiveness, and make improvements to the existing RM process.

REFERENCES

[1] Nikkei Computer, 2003 Survey on Information Actual Condition (2003)

[2] Standish Group International, Inc,"CHAOS Summary 2009," http://www.standishgroup.com, (2009)

[3] K.H. Rose, "A guide to the project management body of knowledge (pmbok guide) fifth edition," Project management journal Vol.44, No.3, (2013)

[4] Project Management Association Japan, P2M Program,Project Management Standard guidebook (2014)

[5] Martin Tomanek and Jan Juricek , "Project RM model based on prince2 and scrum frameworks," International Journal of Software Engineering Applications (IJSEA), Vol.6, No.1 (2015)

[6] B.W. Boehm, "Software risk management: principles and practices," IEEE software, vol.8, no.1, pp.32-41(1991)

[7] R.C.Williams, J.A.Walker, and A.J.Dorofee, "Putting risk management into practice," IEEE software, vol.14, no.3, pp.75-82,(1997)

[8] Kwak, Y.H., Brief history of project management. Chapter 2. In: Carayannis, Kwak, Anbari (Eds.), The Story of Managing Projects: an Interdisciplinary Approach. Quorum Books, Westport, Connecticut.(2003)

[9] Muriana, Cinzia and Vizzini, Giovanni, "Project risk management: A deterministic quantitative technique for assessment and mitigation," International Journal of Project Management, vol. 35, NO.3, PP.320-340, (2017)

[10] Van Slyke, R.M., Monte Carlo methods and the PERT problem. Oper.Res.Vol.11 No.5, PP.839-860.(1963)

[11] Vanhoucke, M., Vandevoorde, S., A simulation and evaluation of earned value metrics to forecast the project duration. J. Oper. Res. Soc. Vol.58 No.10 ,PP.1361-1374, (2007)

[12] Narbaev, T., De Marco, A.D., "An earned schedule-based regression model to improve cost estimate at completion," International Journal Project Management Vol.32 No.6, PP. 1007-1018.(2014)

[13] Acebes, F., Pajares, J., Gal n, J.M., L pez-Paredes, A., "Beyond earned value management: a graphical framework for integrated cost, schedule and risk monitoring," Procedia Society Behav. Science Vol.74, PP.181-189.(2013)

[14] B. Curtis, H. Krasner, N. Iscoe, "A Field Study of the Software Design Process for Large Systems, " Comm. of the ACM, Vol.31, No.11, PP.1268-1287 (1988)

[15] C. Gutwin, K. Schneider, R. Penner, and D. Paquette. "Supporting Group Awareness in Distributed Software Development," Engineering Human Computer Interaction and Interactive Systems, Revised Selected Papers, Springer Verlag, Berlin, pp.383-397.(2005)

[16] J. D. Herbsleb, R. E. Grinter, "Architectures, Coordination, and Distance," Conway's Law and Beyond. IEEE Software, Vol.16, No.5, PP.63-70, (1999)

[17] M. Cataldo, P. Wagstrom, J. Herbsleb, and K. Carley. Identification of Coordination Requirements: Implications for the Design of Collaboration and Awareness Tools. In Conference on Computer Supported Cooperative Work (CSCW'06), Banff, Alberta, Canada, (2006)

[18] S.B. Fonseca, C.R.B de Souza, D.F. Redmiles. Exploring the Relationship between Dependencies and Coordination to Support Global Software Development Projects. In First International Conference on Global Software Engineering, Florian polis, Brazil, pp.243-243.(2006)

[19] J. A. Espinosa, S. A. Slaughter, R. Kraut, and J. D. Herbsleb. "Team knowledge and coordination in geographically distributed software development". Journal of Management Information Systems, Vol.24, No.1,(2007)

[20] Pajares, Javier and Lopez-Paredes, Adolfo, "An extension of the EVM analysis for project monitoring: The Cost Control Index and the Schedule Control Index," International Journal of Project Management, Vol.29,No.5, PP.615-621, (2011)

[21] Priya and Lunge, Harihar S, "Effect of Schedule Performance Index on Cost Estimation in Earned Value Management and Earned Schedule using Weibull, Gamma and Exponential Function,", International Journal of Engineering Science Invention Research & Development; Vol.2, Issue 4, PP.244-250,(2015)

(Received October 4, 2018)
(Revised January 6, 2019)

Akihiro HAYASHI received an MBA and a Ph.D degree from the University of Tsukuba, Tokyo, Japan. His professional career includes International excellent companies: Motorola, NTT, and IBM. He has become a Professor of the University of Shizuoka Institute of Science and Technology in 2018. His research interests include Total Quality Management, Process Improvement, and Quantitative Project Management.

Nobuhiro Kataoka received a Ph.D degree from the University of Tohoku, Japan. He joined Mitsubishi Electric Corp. in 1968 and became a technology director at Information System Technology Center in 1997 and a Professor of Tokai University in 2000-2009. His research interests include Quality Control, Business Process Modeling, Business Modelization Methodologies. He is now a part-time instructor of Tokyo Denki University. He is also a member of IPSJ and a fellow of IEICE.

Yasunobu Kino received a Ph.D. degree from the University of Tsukuba, Japan.He joined IBM Japan Corporation in 1990. He has been in various systems integration projects such as banking system, credit card authorization system, newspaper database, high-speed motion pictures database, and free format OCR. He is an Associate Professor of Faculty of Business Sciences, University of Tsukuba.

Mikio Aoyama received his master and doctor in both engineering from Okayama University and Tokyo Institute of Technology, respectively. From 1986 to 1988, he was visiting scholar at the University of Illinois, USA. In 1995, he became a professor at the Niigata Institute Technology, then a professor at Nanzan University in 2001. His current research interests include cloud computing, requirements engineering, and automotive software engineering.

Invited Paper

Cybersecurity Technologies Essential in the Digital Transformation Era

Kazuhiko Ohkubo

NTT Secure Platform Laboratories, NTT Corporation, Japan
kazuhiko.ookubo.sw@hco.ntt.co.jp

Abstract - An age of digital transformation is currently pressing. New fundamental technologies are penetrating in fields of IoT (Internet of Things) and OT (Operational Technology) in addition to a conventional IT field. Also, a paradigm shift of the ICT environment is producing many advanced economic activities of data use in society. This paper summarizes increasing security risks in such a condition and security technologies we should promptly develop for reducing the risks. Regarding IT security, threat monitoring targets need to be expand to more micro and macro, specifically to endpoint and backbone network. Regarding IoT and OT including critical infrastructure, breakthrough countermeasure functions need to be developed all over functions of the NIST Cybersecurity Framework. Then, we should take into consideration IoT and OT features such as poor computer resources, peculiar industrial protocol, variations of system configuration, etc. Toward development of data use in society, an up-to-date environment needs to be offered by making full use of cryptographic techniques. In other words, both data holders and data users can distribute and handle even privacy and confidential information safely and securely.

Keywords: Cyberattack countermeasure, NIST Cybersecurity Framework, Critical Infrastructure Protection, Secure Computation, Anonymization

1 INTRODUCTION

Digital transformation is the idea that “IT will penetrate every aspect of people’s lives to transform it for the better,” proposed in 2004 by Professor Erik Stolterman of Umea University, Sweden [1]. In the first phase of digitization, work processes were improved through use of IT; in the second phase work is being replaced by IT; and in the third phase work is being transformed seamlessly to IT, and IT to work. In the past, mechanisms such as artificial intelligence and robotics belonged to the world of science fiction, but they are now being realized, partly through innovations in IT technology, to develop a cyberspace society that distinguishes less and less between real and virtual worlds.

Security threats are rapidly escalating. Malware that can act more and more autonomously have begun to appear, cyberattacks are increasingly intelligent. Moreover, Internet of Things (IoT) devices, which are already vulnerable in terms of security, are being connected to the Internet but providing a platform for large-scale DDoS cyberattacks. For these reasons, technologies to counter cyberattacks must continually advance in a game of cat-and-mouse. Also, technologies to counter new anticipated security threats are needed as economic activity develops and ICT environments undergo further paradigm shifts.

On the other hand, the approach of the 2020 Tokyo Olympics and Paralympics is prompting serious concern regarding increasing threats to security in the Operational Technology (OT) domain. This covers control systems and even critical infrastructure, and the concern is whether or not activity surrounding incident prevention and operational responses is adequate if an incident should occur. As such, urgent issues have already become technical development to maintain security for OT domain and strengthening security risk management to improve efficiency of various operational responses.

In addition to "defensive" security measure described above, there is also an increasing need for so-called "offensive" security measure. It enables to ensure safe and secure data utilization businesses that keep pace with the progress of digital transformation. Especially, it will play an important role in light of the enactment of revisions to the personal information protection laws made in May 2017 and the General Data Protection Regulations (GDPR) in May 2018. As such, there is much anticipation for efforts toward risk mitigation using encryption and other information security technologies, and toward new value creation that will contribute to economic revitalization.

Considering these changes in the cyberspace environment and economy, this article discusses threats and security issues being realized in the areas of IT, Non-IT (IoT/OT) including critical infrastructure, and data use in society. Then, it studies on technical development needed to be accelerated for elimination of these issues in the future. Finally, we discuss current conditions and future prospects for security related to AI, which is still advancing today, and anticipates the so-called Singularity, when it is believed AI will exceed human intelligence.

2 CAT-AND-MOUSE IT SECURITY

For the PyeongChang Winter Olympics in February 2018, targeted attacks on Olympics-related organizations occurred from the beginning of the year, and several incidents caused damage around the time of the opening ceremony. Specifically, the public Web site, the press center network connection and the stadium wireless LAN all went down temporarily, and the drones intended for use during the opening ceremony did not fly. We conducted an independent investigation, acquiring and analyzing malware used in the cyberattacks, called Olympic Destroyer. As a result, we found that the malware caused destructive activity in PCs and other devices, and that the attack was clearly intended to disrupt the event. Besides, recent cyberattacks

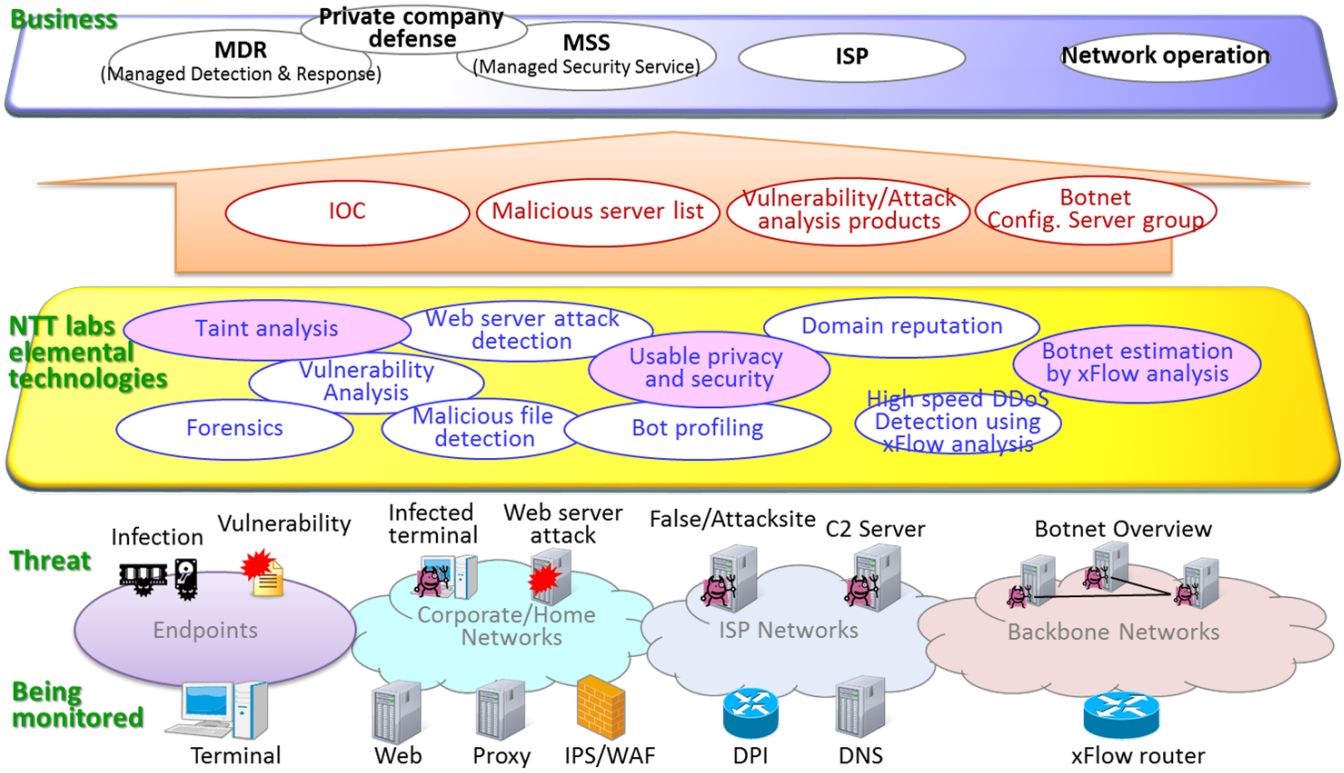


Figure 1: Cyberattack countermeasures in a game of cat-and-mouse.

are becoming not only more ingenious but also more increasing in scale by utilizing many devices infected with malware to conduct DDoS attacks.

Cyberattacks are getting smarter and increasing in scale in these ways, so the scope of monitoring needs to expand in order to oppose them. Conventionally, corporate, home and ISP networks were monitored, but this must expand to include both micro and macro perspectives, from endpoints to backbone networks (Fig. 1). Regarding endpoints, advanced Indicators of Compromise (IOC) are generated by using technologies such as taint analysis to precisely analyze malware behavior. Such IOC can be effectively used in Managed Detection and Response (MDR) and other products. On backbone networks, analysis of large volumes of flow data can reveal the overall structure (Herder, C2 server and bot terminals) of a botnet, and provide clues to appropriate countermeasures.

Although security technology developed in the past focused mainly on system and network security, a technical area called "Usable privacy & security" is recently increasingly getting attention in the world. It takes the perspective that it was the user that was ultimately deceived, and focuses on human-computer interaction. Specifically, it attempts to improve security by identifying causal gaps due to user unawareness or inappropriate action, and making system improvements accordingly. Such gaps in privacy and security are also leading to promising advances in honeypots that mimic user behavior and intelligent technologies for security experts.

3 SECURITY IN THE NON-IT FIELDS OF IOT/OT

The first time that Mirai, a prototypical malware infecting IoT devices, was used for a large-scale attack was a DDOS attack to the "KrebsOnSecurity" security blog in September 2016. It was the largest that had been seen at the time, at a reported 620 Gbps. The Mirai source code was then immediately released by someone called Anna-senpai, resulting in the creation of many variants, most of which were used in a stream of other large scale DDoS attacks. Figure 2 shows the Mirai attack mechanism that we clarified by downloading and analyzing its source code. According to this, we can find not only the attack mechanism is complicated but also each IoT device is extremely vulnerable. Specifically, high-speed telnet port scan and brute-force attack can be easily done. In these incidents, most of the owners were not aware that their IoT devices were being used in large-scale DDoS attacks. On the contrary, IoT device owners are being embarrassed by ransom-ware type variations. The ability to infect IoT devices with ransom-ware, rendering them inoperable if a ransom was not paid (by bitcoin for example), had already been verified in the laboratory. So, it is just a matter of time that this occurs as well.

In March 2018, as also reported in newspapers, it was discovered that the administrator screens of several hundred routers at a telecommunications operator were visible from the Internet. This is something generally true for IoT devices, but it is assumed that in all cases, the fundamental issue was

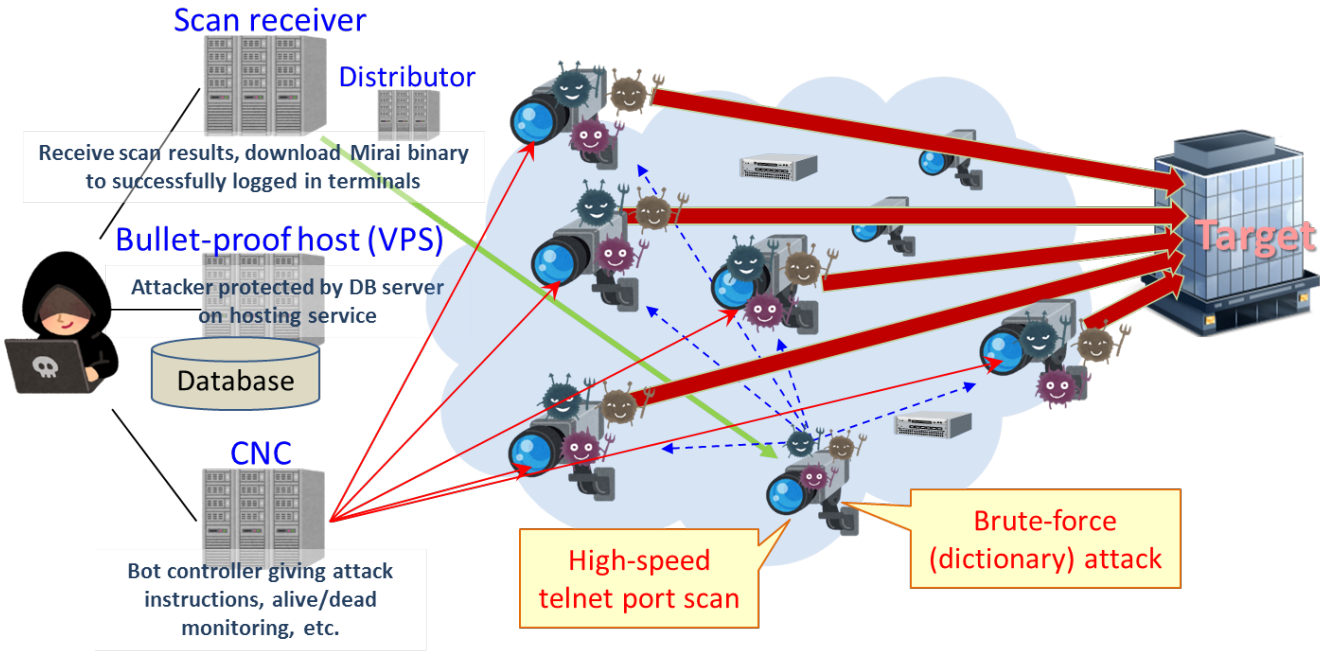


Figure 2: Mirai attack mechanism.

that the Web-UIs were open to external access for any of the following reasons.

- Default ID/PWDs were extremely simple
- Operation possible without changing default ID/PWD
- The same default ID/PWD is used for all units of a given product or all products from a given vendor
- Online manuals giving default ID/PWD can be seen online by anyone

To address the issue, the Ministry of Internal Affairs and Communications in Japan has revised regulations governing communications operators to begin implementing necessary measures within 2018 [2]. As part of this, National Institute of Information and Communications Technology (NICT) Act is working to be revised. Specifically, it will augment NICT's duties with details such as adding a five-year limited survey of IoT devices with inadequate password settings. The revision and enactment of laws are scheduled to complete in fiscal year 2018.

From a technical development point of view, because IoT devices are limited in computing resources of CPU, memory, disk space, battery and the like, existing IT security functions such as anti-virus software cannot be used. Thus, a new range of security technologies for IoT devices covering authentication/authorization, configuration management, and detection and response must be established from scratch. Such necessary technologies can also be classified into functional elements of the USA National Institute of Standards and Technology (NIST) Cybersecurity Framework [3]. The Framework enables organizations regardless of size, degree of cybersecurity risk, or cybersecurity sophistication to apply the principles and best practices of risk management to improve security and resilience. The Framework provides a common organizing structure for multiple approaches to cybersecurity by assembling standards, guidelines, and practices that are working effectively today.

For authentication/authorization, an example would be a next-generation authentication technology not requiring password management at the server (Fi. 3). Such a technology would provide some secret information to the device from an initial registration server when the client is first initialized. It would be used together with a simple, device-specific ID, like the PIN number used with a cash card, to implement authentication. This would allow operation without managing passwords of individual IoT devices, and cost reduction of issuing and using certificates needed for authentication.

Regarding configuration management, detection and response, when various IoT devices are connected under a gateway, devices can be identified or estimated accurately. Then, the configuration must be discovered even in LAN environments with severe operating conditions. This can be done by analyzing the output characteristics of commonly used ARP frames and using noise cancelation. IoT device specific information can then be used to discover devices with vulnerabilities. Graph theory and other techniques can also be used to detect traffic anomalies excluded from a white list of usual communication counterparts. This enables to classify abnormal communication as a part of an attack or otherwise, and handle it with a communication control alert, quarantine, or other means.

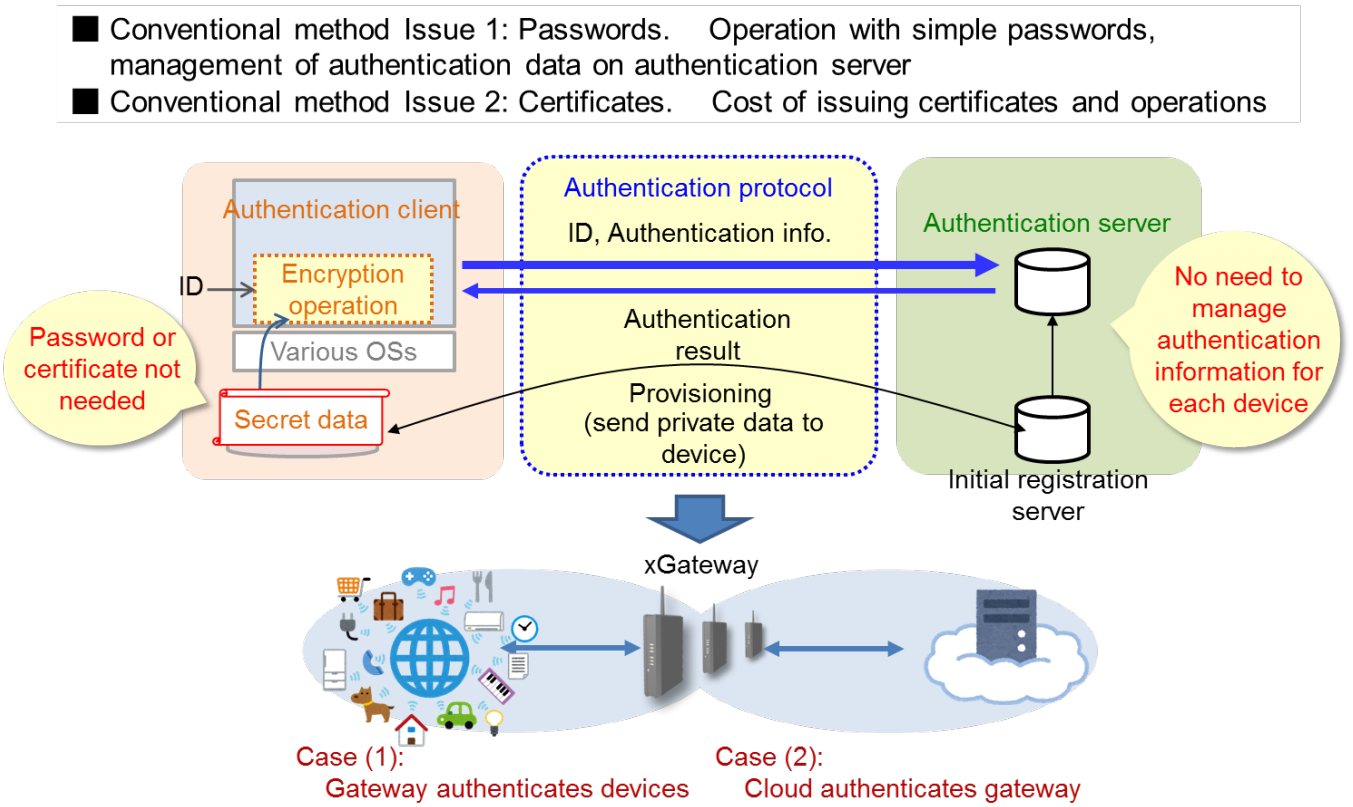


Figure 3: Next generation authentication technology for IoT.

Similarly, security technology for OT must also be re-established from the start, and unique aspects particular to this domain, such as industrial protocols, must also be handled. InteRSePT® is composed of "Real-time detection/handling" and "Security integration management" through a joint development between MHI, Mitsubishi Heavy Industry, and NTT [4]. Sensor and other data on the network is comprehensively monitored by InteRSePT to detect malicious cyberattacks using control commands that were difficult to deal with using earlier technology. Security rules can be changed in real time for each operational state of the devices, to detect anomalies quickly, handle unknown cyberattacks and maintain system availability.

4 SECURITY ASPECTS OF PROTECTING CRITICAL INFRASTRUCTURE

From 2009 into 2010, a malware called Stuxnet infiltrated nuclear facilities in Iran causing real damage and becoming the first known cyberattack on critical infrastructure. The fact that industrial control systems not connected to the Internet could be infected through USB memory was a major shock at the time.

When considering the increasing cyber-risks for critical infrastructure, important points include the followings.

(1) Characteristics of being large-scale and involving complex linked systems
(2) Changes in the environment toward application of general purpose, open, and new technologies

Regarding the former (1), it is not unusual for infrastructure facilities to have thousands of server devices, and tens to hundreds of thousands of control devices. If a cyberattack is successful on even one of these locations, the effects could be widespread. Thus, technology is needed to continually check that components are authentic and have not been illicitly infiltrated or modified, to prevent abnormal operation as shown in Fig. 4. Authenticity checking technology builds a chain of trust (trust reference points) and enables to reliably detect any system falsification occurring on a large-scale system, system-wide and from startup through operation.

The authenticity checking resembles a technique used in an IC (Integrated Circuit) passport. In other words, an IC passport has a plastic card with a contactless IC chip built into the center of the passport booklet. It stores basic passport information including the passport holder's name, nationality, birth date, and passport number, as well as a facial image (exactly digest data of facial image) read from the photo in the PDF of the passport application. At passport control, it is very easy for an officer to discover foreigners that substituted the facial image by comparing the digest data such as "hash value." It is just a concept to introduce

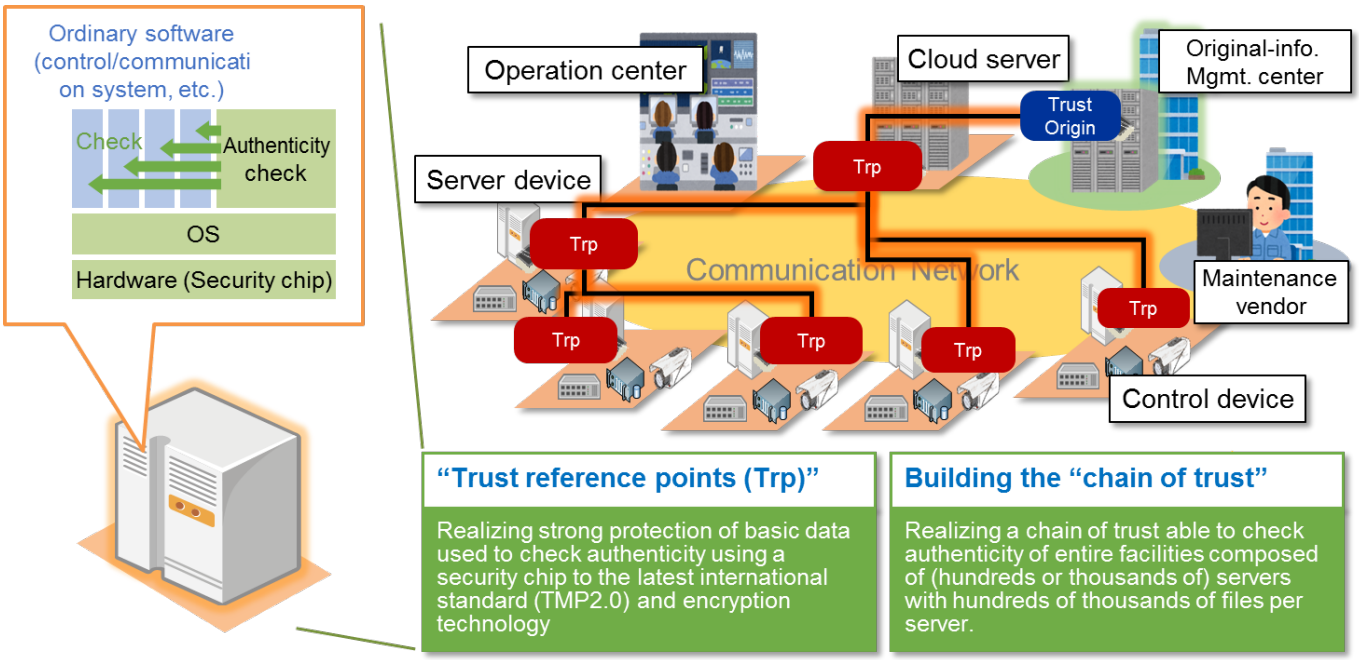


Figure 4: Detecting system falsification with authenticity checking technology.

such a mechanism into individual devices that are components of a critical infrastructure facility.

Regarding the latter (2), Internet technologies and open source software such as Linux continue to be adopted, making it easier to obtain vulnerability and other information needed for attacks. As such, a major assumption is that, for devices and networks where the authenticity checking technology cannot be built-in, there is a need for bolt-on technology able to monitor and analyze system behavior for anomalies.

Behavior monitoring and analysis technology can adapt automatically to diversifying devices and handle unknown attacks by using AI technology (unsupervised deep learning). Also, it can handle particularities of control communication, with signals from multiple devices having unique packets overlapping within several milliseconds.

Parts of both the authenticity checking technology and behavior monitoring and analysis technology are currently under development by the New Energy and Industrial Technology Development Organization (NEDO). It is done under the Council for Science, Technology and Innovation Strategic Innovation creation Program (SIP) called "Cyber-security for Critical Infrastructure". Please see the related Web site for details [5] [6].

New 5G technologies are also under development for implementation and commercialization in 2019. So, work to identify risks brought by the spread of these new technologies and to study security measures to deal with them will become increasingly important in the future. Specifically, increased risk could come from more powerful attacks utilizing characteristics of 5G networks, including higher bandwidth, ultra-low latency, and more simultaneous connections. New types of cyberattack may also come from diversification in the use and forms of IoT devices, increasing dependence of infrastructure, and even from attacks on networking devices including devices at telecommunications providers. As such, conventional network security architectures focusing on protecting voice and data must be supplanted. Therefore, a new security architecture for 5G should be studied urgently, emphasizing the following vital perspectives.

1. Network, service, and hybrid user authentication
2. Virtual network slice security management
3. Network countermeasures for large-scale DDoS attacks
4. Traffic monitoring and anomaly detection, including on wireless segments
5. Protection of private information (ID, location data, personal content, etc.)

5 DATA UTILIZATION IN SOCIETY

According to revisions made to the personal information protection law and enacted in May 2017, personal data processed to create "anonymized data" can be provided to third parties without agreement from the people involved as shown in Fig. 5. Anonymized data is defined as information regarding individuals that has been processed such that particular individuals cannot be identified from the processed data. In creating such data, all regulations 1 to 5 stipulated in Article 19 of the enforcement regulations [7] must be met [8]. Rules No. 1 to 4 are very easy to be done because of data deletion regarding name, biometric data, ID, very expensive purchase, etc. On the other hand, rule No. 5 is very difficult to be done in spite of such an exemplification as "If there is data regarding elementary

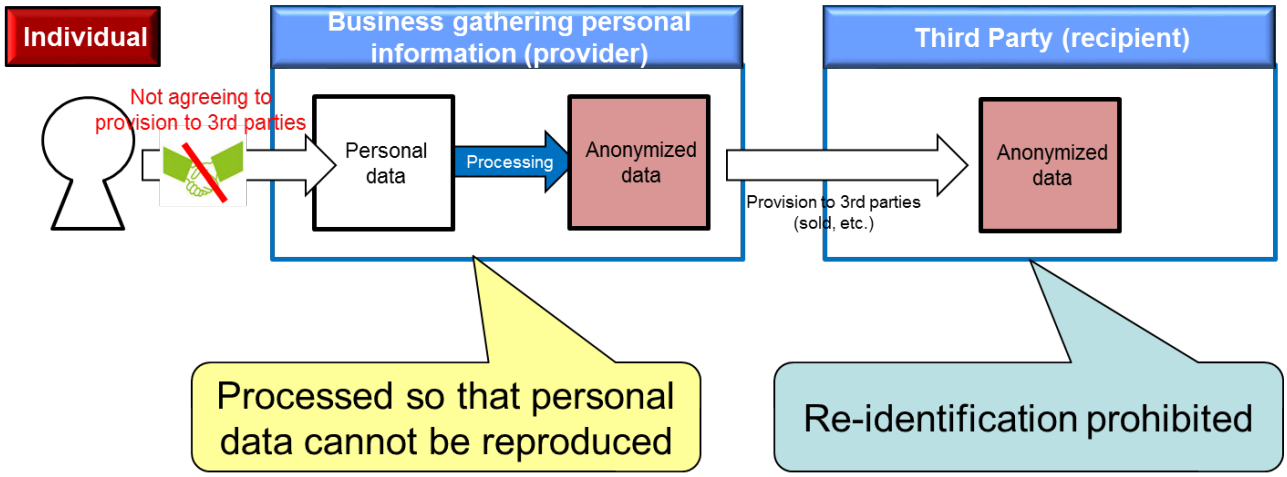


Figure 5: Outline of the revised Personal Information Protection Act.

school students over 170 cm tall, replace it with “150 cm or taller".

1. Delete descriptions which can identify a specific Individual
2. Delete personal identification codes
3. Delete linkage codes which link personal information and obtained information
4. Delete idiosyncratic descriptions
5. Take appropriate action considering the properties and differences between descriptions in personal information

With k-anonymization, which is a typical advanced anonymization technique, k-anonymity, which is an index of safety, is achieved through data generalization; data is processed such that it cannot be narrowed down to k or fewer persons with the same information. However, in doing so it is difficult to maintain both safety and usefulness of the data. In contrast, Pk-anonymization [9] is considered relatively more effective, because it can achieve safety equivalent to k-anonymization while preserving usefulness. As shown in Fig. 6, the method executes data randomization, in other words, addresses are replaced keeping city level and ages are also replaced keeping at 1 position different from data generalization by k-anonymization.

There is also a need around the world to use data without releasing it externally, even in anonymized form. “Secure Computation” involves processing data in its encrypted form and can be useful for such cases. There are many schemes for secure computation. However, schemes based on secret sharing [10] [11], which is an ISO standard, are the most practical from the perspectives of the definition of safety,

Name	Address	Sex	Age	Occupation
Sato	Shinjuku, Tokyo	M	45	Company Employee
Suzuki	Mitaka, Tokyo	M	41	Company Employee
Abe	Shinjuku, Tokyo	F	37	Homemaker
Nagasawa	Shinagawa, Tokyo	F	35	Homemaker
Yamamoto	Funabashi, Chiba	M	51	Self-employed
Kobayashi	Chiba City, Chiba	M	57	Self-employed
Uchida	Kashiwashi, Chiba	M	59	Self-employed

Random-ization

Name	Address	Sex	Age	Occupation
	Shinjuku, Tokyo	M	57	Company Employee
	Mitaka, Tokyo	M	41	Self-employed
	Funabashi, Chiba	F	37	Homemaker
	Shinagawa, Tokyo	M	35	Homemaker
	Shinjuku, Tokyo	M	51	Company Employee
	Chiba City, Chiba	M	45	Self-employed
	Kashiwashi, Chiba	F	59	Self-employed

Maintains safety equivalent to k-anonymization and preserves data usability

Figure 6: Pk-anonymization.

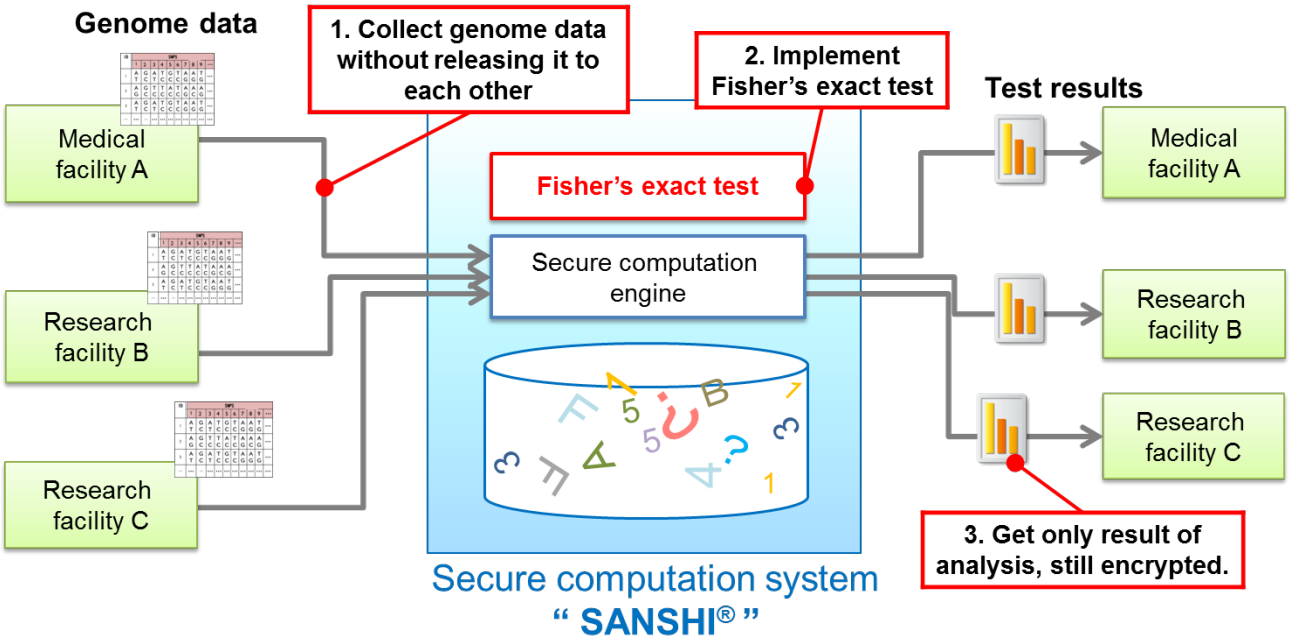


Figure 7: Secure computation system: "SANSHI". (算師®)

general purpose computations, sensible performance and international standardization. We hope use of this technology will spread in the future. Incidentally, NTT together with Tohoku Medical Megabank Organization (ToMMo) have used secure computation to implemented Fisher's exact test, to analyze the relationships between human DNA variations and diseases [12]. This is the first such implementation in the industry (Fig. 7).

6 SECURITY FOR THE SINGULARITY (2045 PROBLEM)

Singularity, also called Kurzweil's law of accelerating change, suggests that by 2045 as follows; a $1,000 computer will have performance of approximately 10 peta FLOPs, which is ten billion times that of the human brain and a sufficient base for AI to reach a technical singularity. Automated cyberattacks using AI are already appearing, advanced hacking using AI is likely to become mainstream. As well, it will become necessary to use automated technologies with AI on the defensive side. Actually at the world's largest hacking contest in DEFCON 2016, 7 computers automatically hacked each other's computer. So, it is no exaggeration to say that AI hacking has already begun. In doing so, attacks to machine learning have also been identified as an issue. Examples of attacks are to create input that induces false recognition, to contaminate classifier training data, and to steal the classifier itself by submitting queries to the classifier [13].

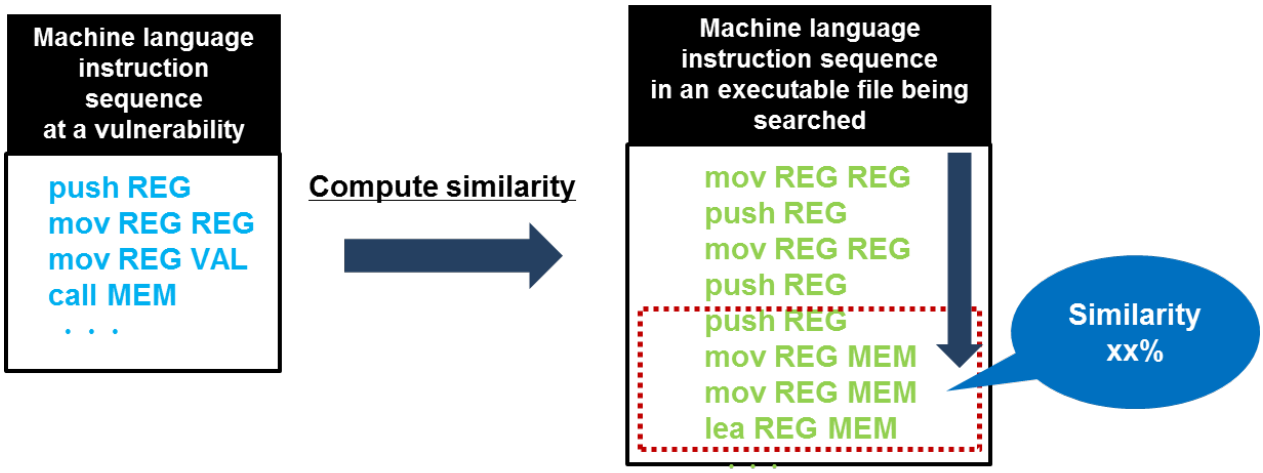


Figure 8: AI hacking-related technology to identify vulnerable points.

To deal with these, there is an urgent need to establish technologies able to detect vulnerabilities in executable binaries and to detect conditions that trigger an attack using symbolic execution. Regarding vulnerability detection, for instance, comparing an existing vulnerable binary code and a target binary code enables to identify vulnerable points by computing similarity as shown in Fig. 8.

Technical development of AI will bring great change and diversification in human thought/behavior and assumptions regarding societal structures. Therefore, research on legal systems, which function as societal standards, is also becoming crucial. The following steps need to be taken for a smooth transition from AI development to societal implementation.

1. Assuming application of AI, anticipate potential effects on people, society, and industry, and real relationships among them
2. Check current laws with knowledge of AI and analyze individual concerns
3. Propose a new legal system for the AI era, for future legislation and policy

Related activity is appearing in Japan [14], and related joint research is being actively developed at the RIKEN Center for Advanced Intelligence Project (AIP) and NTT Laboratories.

7 CONCLUSION

This article has given a comprehensive outline of security technologies essential in future technical development, mainly to eliminate threats and security issues anticipated with the arrival of digital transformation era. It has discussed, from a technological point of view, both offensive and defensive security perspectives. Then, it has given consideration to individual functional elements of the USA National Institute of Standards and Technology (NIST) Cyber Security Framework: Identify, Protect, Detect, and Respond. Moreover, it stated that perspectives of not only technology but also legal system are necessary for efficient security-risk management and effective countermeasures implementation.

REFERENCES

[1] Digital Transformation https://en.wikipedia.org/wiki/Digital_transformation
[2] Partial revision of NICT Regulations http://www.soumu.go.jp/main_content/000536856.pdf
[3] Framework for improving critical infrastructure cybersecurity http://www.nist.gov/cyberframework/upload/cybersecurity-framework-021214.pdf
[4] InteRSePT®: A Cybersecurity technology realizing safe and secure operation of control systems enters the market http://www.ntt.co.jp/news2018/1804e/180425b.html
[5] What is the Cross-ministerial Strategic Innovation Promotion Program?http://www8.cao.go.jp/cstp/panhu/sip_english/5-8.pdf
[6] Secure Architecture for Critical Infrastructure https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201705fa2.html
[7] Enforcement Rules for the Act on the Protection of Personal Information (Tentative translation) https://www.ppc.go.jp/files/pdf/PPC_rules.pdf
[8] R, Osumi, K. Takahashi: “Personal Data Anonymization and Use,” Seibunsha (Japanese).
[9] Dai Ikarashi, Ryo Kikuchi, Koji Chida, Katsumi Takahashi: "k-Anonymous Microdata Release via Post Randomisation Method," International Workshop on Security (IWSEC), 2015
[10] NTT Secret Sharing technology Selected as First International Standard for Secret Sharing Technology (Japanese) http://www.ntt.co.jp/news2017/1710/171023a.html
[11] ISO/IEC 19592-2 Information technology -- Security techniques -- Secret sharing -- Part 2: Fundamental mechanisms
[12] Koki Hamada, Satoshi Hasegawa, Kazuharu Misawa, Koji Chida, Soichi Ogishima, and Masao Nagasaki: "Privacy-Preserving Fisher's Exact Test for Genome-Wide Association Study," International Workshop on Genome Privacy and Security (GenoPri), 2017.
[13] David Wagner on Adversarial Machine Learning at ACM CCS’17 https://syncedreview.com/2017/11/07/david-wagner-on-adversarial-machine-learning-at-acm-ccs17/
[14] Ministry of Internal Affairs and Communications, “AI Network Society Promotion Council,” http://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/

(Received October 8, 2018)

Kazuhiko Ohkubo is a vice president and the head of NTT Secure Platform Laboratories. He received his B.S. in information engineering from the University of Tsukuba in 1987 and M.S. in electrical engineering from the University of Tokyo in 1989. He also earned his M.S. degree in management of technology from the MIT Sloan School of Management, USA in 2000. He is a member of IEICE and IEEE.

Industrial Paper

Training Data Generation Method for Deep Learning by Utilizing Computer Graphics

Tsukasa Kudo†, Ren Takimoto†, and Tenma Kawanaka‡

†Faculty of Informatics, Shizuoka Institute of Science and Technology, Japan
‡Shizutetsu Information Center Corpration, Japan
kudo.tsukasa@sist.ac.jp

Abstract - In image recognition, deep learning has enabled innovative improvement in accuracy. As a result, its application has spread to various fields. However, in order to conduct deep learning, it is necessary to accumulate a large number of training data. And, this often becomes the obstacles to applying the deep learning to actual business systems. On the other hand, for computer graphics (CG), various tools have been developed and provided. And, currently, we can not only create a CG image easily but also execute various CG operations automatically by program control. So, in this study, we propose a method to generate the images of training data automatically by using CG. For example, in inventory management of parts, target objects are composed only of inventory shelves and parts with heavy but simple shapes. That is, although it is difficult to accumulate a large number of their actual images, it is expected that these images can be easily generated by using CG. Furthermore, we create CG images for the experiments and automatically generate training data to conduct deep learning for inventory quantities. Then, we conduct experiments to evaluate the estimation accuracy with respect to the inventory quantity shown in the actual inventory photograph. Through this experiment, we show this method is effective in some fields where it is difficult to accumulate a large number of the actual images for the deep learning.

Keywords: Deep learning, Computer grapahics, CG, Inventory management system, Stocktaking, Convolutional neural network

1 INTRODUCTION

Recently, in the field of pattern recognition such as speech recognition and image recognition, the effectiveness of deep learning has been confirmed [17], [13]. As a result, its applications are rapidly spreading to various fields [4], [8]. And, by applying it, recognition accuracy has been rapidly improved; especially in image recognition, the case of achieving accuracy exceeding human vision has also been reported [7], [9], [20]. As one of the reasons for such rapidly spreading, it can be pointed out that a large number of data necessary for learning became to be prepared easily. That is, with the progress of the Internet of Things (IoT), a large number of data such as images and movies are made public on the Internet and data collection also became facilitated [1], [5].

However, there are some fields where the collection of training data is not easy. For example, in the manufacturing factory of mechanical products which our laboratory supports its production management system improvement, the stocktaking of parts inventory is a heavy workload. Especially, since parts quantity in the bulk container cannot be counted from the outside, they must be taken out to count. So, in our previous study, we proposed a method to automatically discriminate the inventory satisfaction by image recognition utilizing deep learning, then constructed and evaluated its prototype. As a result, we found this method is useful, and high reliability can be obtained by comparing with the theoretical inventory obtained by the production management system [12].

For these evaluations, we prepared 1,600 image data for each part. However, in the actual factory, the parts are delivered to the assembly field from parts shelves collectively: by each product lot, namely product manufacturing unit, or by each order composed of several products. So, the number of times of variation of the shelves is comparatively small. For example, in the case where its variation occurs once a day, about 250 images are obtained a year. That is, to accumulate 1,600 images, it takes more than 6 years. In addition, parts are stored in each inventory shelf, and their kinds extend to thousands. And, since most parts are heavy, it is not practical to deliberately change the situation of the inventory shelves by hand so many times. For these reasons, to prepare the training data efficiently has become the problem in applying the deep learning to the image recognition for the actual inventory satisfaction discrimination.

Here, each image of an inventory shelf is composed of only two types of objects: the inventory shelf itself and the part. And, the parts are placed on the shelf according to a certain rule. For example, in the case of storing relatively small parts in a bulk container, the state of the inventory shelf can be composed by piling up the parts randomly from the bottom of the container. Furthermore, the parts have simple shapes such as nuts and bolts, and there are relatively few types of materials for parts. This suggests that a large number of various image data of each inventory shelf can be generated by the computer graphics (CG) efficiently.

The motivation of this study is to show there are fields as follows: it is difficult to accumulate a large number of training data composed of photographs of the actual objects (hereinafter, 'actual images') for the deep learning; but, it is easy to accumulate the training data by utilizing CG. The target of this study is the above-mentioned inventory satisfaction discrimination method in our previous study, which was carried

Figure 1: Inventory shelves of bulk container.

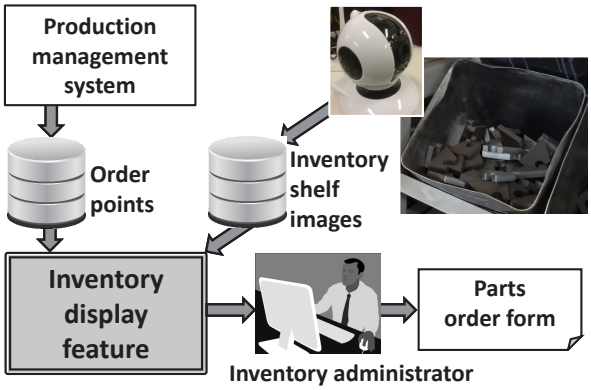


Figure 2: Inventory management utilizing images.

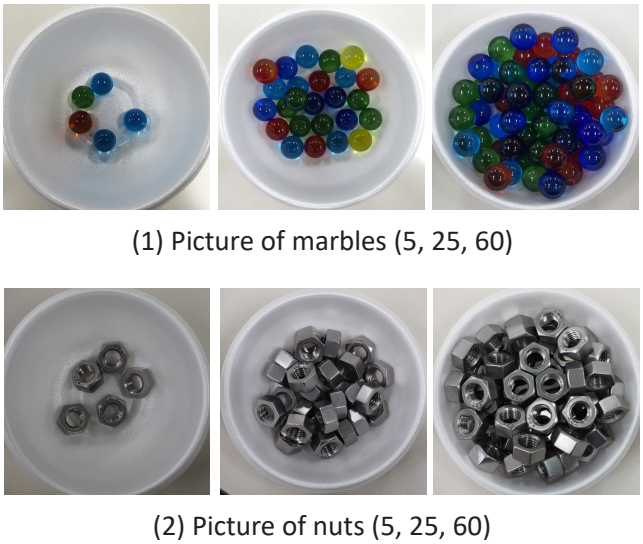


Figure 3: Picture example of experimental objects.

out by using only the actual images for the training data. In this paper, we show the training data generated by CG tool can complement the actual images, that is, the lack of the actual images can be supplemented with the CG images.

Concretely, this paper shows the following four points. The first is the relationship between CG image factors and recognition accuracy, and we show that comprehensive improvement of CG image is necessary to improve the accuracy. The second is the case where both of actual images and CG images are used, and we show the accuracy can be improved by adding some actual images into CG images. The third is the case for different parts, and we show there is a similar tendency with respect to the above-mentioned points. The fourth is the evaluation of man-hours to create CG images of parts on the premise of using the same material and simple shape, and we show it is much more efficient than collecting a large number of actual images.

The remainder of this paper is organized as follows. Section 2 shows the related works and the problem to create the training data, and we propose the training data generation method with CG for the stocktaking in Section 3. Section 4 shows the implementation of this method, and Section 5 shows the experiments and evaluations with the training data generated by this method. We discuss the evaluation results in Section 6 and conclude this paper in Section 7.

2 RELATED WORKS AND PROBLEM

In this section, we explain the background and related works of this study. Our laboratory supports a factory, which manufactures mechanical products, to introduce and operate its production management system. Since thousands of parts in various shapes are stored in each inventory shelf in the factory, the workload required for stocktaking of inventory is a serious problem. In particular, in the case where parts are stored in the bulk container as shown in Fig. 1, they cannot be counted from the outside. So, since it is necessary to take out the parts from the container and count them up, it is a serious factor increasing man-hours.

On the other hand, with the progress of the Internet of Things (IoT), various sensors such as surveillance cameras are controlled remotely, and their data is accumulated and analyzed in the server. As a result, since so various and enormous data has been stored in the database, it has become difficult to deal with such data with conventional relational databases. So, various NoSQL databases have been put to practical use to manipulate such a data efficiently [16]. For example, MongoDB is a kind of document-oriented NoSQL database and provides the GridFS interface to manipulate such data efficiently in the distributed environment [2]. That is, now a day, large capacity of image and video data have become to be easily handled.

Due to such a technical background, we proposed an inventory satisfaction discrimination method using the inventory shelf images shown in Fig. 2 [14]. Here, the ordering point quantity is determined for each part in the inventory shelves by the production management system, and inventory is replenished when its inventory quantity falls below this ordering point [19]. Therefore, to discriminate visually that the inventory of each part satisfies this ordering point quantity is more efficient than performing the stocktaking of the inventory. For example, in the case of the inventory shelf on the upper right of Fig. 1, while it is difficult to grasp its exact quantity, it is relatively easy to discriminate that there are ten or more parts.

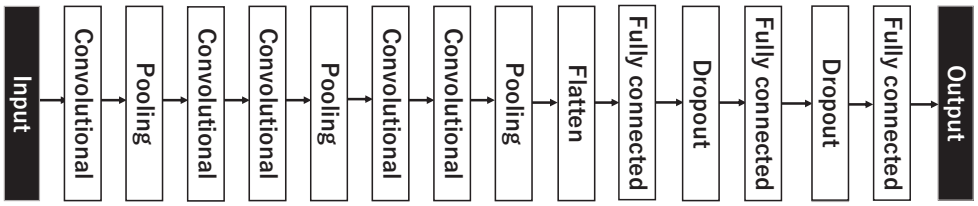


Figure 4: Construction of deep convolutional neural network.

And, in the case where the discrimination was difficult, by replenishing the part from the viewpoint of safety, it became not necessary to count the inventory quantity. As a result, we showed that the efficiency of the inventory management could be achieved by showing both the production plan data and current images of the inventory shelves to the inventory administrator as shown in Fig. 2. However, even by this method, some problems about the workload of the inventory manager remained: he had to check many inventory shelves one by one; especially, in the case where a large number of parts were stored in the bulk container, it took time for the discrimination.

On the other hand, currently, the accuracy of image recognition is rapidly improving by utilizing deep learning. For example, in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), competition in the algorithms for object detection and image classification of large-scale is done. And, after the deep learning was applied in 2012, the recognition rate has been greatly improved. Moreover, it exceeded 5.1% of human recognition rate in 2015 [20]. Along with the improvement in recognition rate, the application of deep learning to image recognition is spreading in various fields such as face authentication, medical image diagnosis, plant disease detection, and so on [4], [8], [15].

So, we conceived to apply the deep learning to the inventory satisfaction discrimination utilizing image recognition. And, as a feasibility study, we conducted the experiment to recognize the classified quantity of the objects by the supervised deep learning as following [12]. We constructed the deep convolution neural network (deep CNN) and performed the deep learning; we evaluated its accuracy of the quantity estimation by using the test data of images prepared separately. For the target objects, we used the marbles and nuts, which is easy to create the training data. And, we classified each of them from 5 to 80 every 5 with the label of their quantity.

Figure 3 shows the image examples of training data in the case of 5, 25 and 60. We prepared 100 image data for each class, that is, the total is 1,600 for each of marble and nut. Then, we padded 400 images by up/down and right/left inversion, and 90 and 270-degree rotation. As a result, we created 500 image data for each class, that is, the total was 8,000 for each of marble and nut. Next, we separated them into 6,000 training data and 2,000 test data, then we separated 600 data from the training data as the verification data. Ultimately, with 5,400 training data, we conducted supervised deep learning with the above-mentioned labels. Figure 4 shows the composition of deep CNN for this deep learning.

After that, we evaluated the accuracy of the quantity estimation in this deep CNN by using the above-mentioned test data. Firstly, we conducted the comparative evaluations between the deep CNN and human vision by using marbles. As a result, we found that although the human vision was able to accurately estimate the quantity in the case where it was small, the deep CNN's accuracy was higher in the case where the quantity was equal or more than 20. This experiment corresponded to the inventory shelf of the bulk container shown in Fig. 1. For example, it was difficult for humans to estimate the quantity of the image of 60 in Fig. 3 in a short time. Furthermore, we evaluated the distribution of the estimated quantity, and we concluded that it could be applied to actual inventory management systems by taking the following countermeasures: the safety stock should be increased to permit the error; the estimated quantity should be compared with the logical inventory quantity calculated by the production management system to detect the error.

In this experiment, since we treated small marbles and nuts shown in Fig. 3, it was easy to obtain a large number of photo to create the training data by shuffling these objects in the bowl at every time with the human hand. However, even with such objects, we took more than one week to take all the photos. Furthermore, in the actual factory, the bulky and heavy parts shown in Fig. 1 are treated. In addition, since it is in operation, we cannot hinder the workers. That is, there is a problem that it is difficult to accumulate a large number of training data for applying the deep learning.

Here, Generative Advisory Network (GAN) has been proposed to generate similar images of actual images, which utilizes deep learning [6], [9]. It consists of the following two parts: the generator network creates images that are intended to come from the same distribution as the training data; and, the discriminator network examines the images to determine whether they are real or fake. And, through training, the former becomes to create fake images that are harder to be discriminated from the real images. In this method, though various similar images can be automatically created, there is a problem that it requires costs of the training and network adjustments.

Generally, to apply the deep learning, accumulating a large number of training data is an important factor to improve its performance, and training data is collected in various ways in each application field. On the other hand, there are application fields where collecting the training data is difficult like this case. Therefore, it is considered effective to develop an

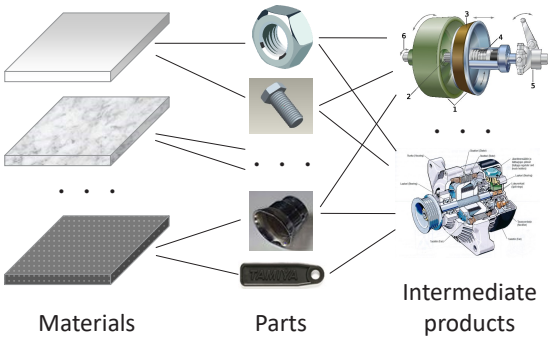


Figure 5: Relationship between materials and parts.

Figure 6: Image example of experimental objects.

efficient preparing method of the training data in such a field.

3 PROPOSAL OF TRAINING DATA GENERATION METHOD UTILIZING CG

As a solution to the problem in the fields where the accumulation of a large number of training data is difficult, we propose a training data generation method by utilizing CG. Considering the characteristics of the inventory shelf, usually, only one kind of parts are stored in one shelf. That is, from the viewpoint of CG, it is possible to construct the state of the inventory shelf with only two objects, one inventory shelf and one part. Furthermore, it is not necessary to consider the deformation of each object. That is, by piling up the part objects randomly in the inventory shelf object, the state of the inventory shelf can be constructed virtually.

In addition, as shown in Fig. 5, in the case of machine parts, although the types of parts are several thousand, the number of types of material is relatively few such as iron, stainless steel and so on. For example, it is less than 20 in the target factory. Also, each part has a simple shape as shown in Fig. 1, and these are combined to produce the 'intermediate products' in Fig. 5; and, the final product is assembled by using them. Therefore, from the viewpoint of CG modeling, namely creating CG objects of parts, only a simple shape processing is the main work. Incidentally, we use the same objects as our previous study, namely marbles and nuts shown in Fig. 3. And, we perform the comparative evaluations of the accuracy between the cases of utilizing the CG image and actual image in deep learning.

Basically, to construct such inventory shelves by CG, the four factors should be considered: the shape of the objects, the material of the objects, the illumination as the environment, and the placement of each object on each inventory shelf. Firstly, the shapes of objects can be realistically created by using CG modeling tools, based on the measuring data of the actual inventory shelf and part. Secondly, the material expresses the textures of these objects, and it can be also added by the CG modeling tools. However, it is necessary to adjust the material by not only human vision, but also checking the influence to the deep learning. Thirdly, the illumination needs to be determined based on the actual factory illumination environment, and it can be added in the same way as the material. These three factors are static with respect to each inventory shelf. So, it can be used repeatedly after once created.

Fourthly, the placement of the part objects is the most important factor to create a large number of training data, and the following three requirements should be considered. The first is a physical requirement, for example, it is necessary that the placement in CG does not collapse even if it is actually placed. The second is the realization of random placement. That is, in order to accumulate a large number of training data for images of the same part, different arrangements must be made for each image, even for the same number of parts. The third is the rule of placement. For example, the parts in the second container from the left of the uppermost shelf of Fig. 1 are the one like plates, and they piled for every 4.

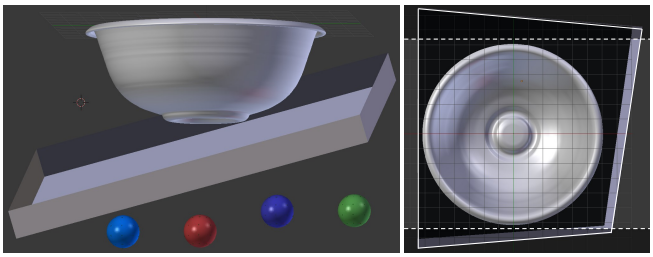
Regarding such a placement, various kinds of CG tools are provided. And, many of them can be controlled automatically by the programming language and provide the feature of physical simulation such as free-fall by gravity. Therefore, it is possible to create the various placement state similar to the real world by executing the physics simulation after randomly placing parts by using programming languages.

For example, we show the case to generate the marble training data shown in Fig. 3 by CG tool. Firstly, as shown in (a) of Fig. 6, we create the bowl and marble objects, then place the marble objects randomly above the bowl. Next, the marble objects are dropped in the direction of the arrow shown in (a) of Fig. 6 by using the physical simulation of free-fall by gravity. After falling in the bowl, they converge to the natural state shown in (b) by the control of the physical simulation. And, by saving the rendering image of this result, the image shown in Fig. 3 can be obtained. By repeating these processes by using the programming language, various large number of training data can be created automatically.

As described above, it is expected that a large number of training data can be generated efficiently by using CG tools in the following case: the target images are composed of a small number of objects; and, a large number of different image data can be created according to a certain procedure. As a result, it is considered that the training data can be generated with CG images as the substitute for actual images in a certain field, where it is difficult to accumulate the training data by actual images.

```
> blender --background --python Marble_basic.py
```

Figure 7: Batch file to control Blender by Python.



(a) Initial placement of objects in Blender (b) Camera view

Figure 8: Initial placement of objects in Blender.

4 IMPLEMENTATION OF TRAINING DATA GENERATOR

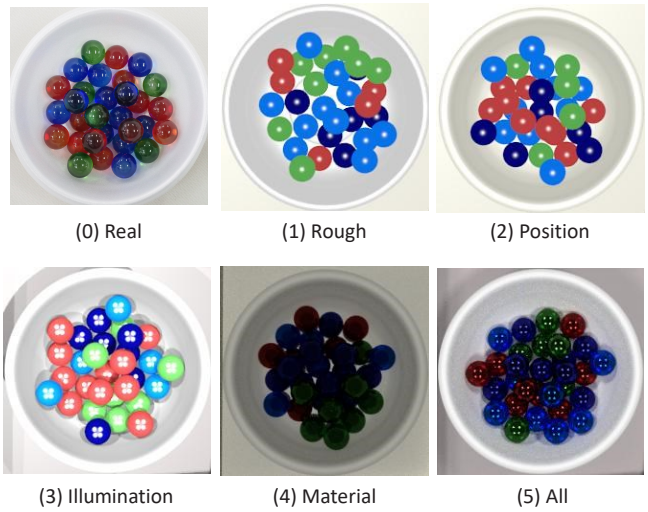
In this study, we used 3DCD creation software Blender 2.79 [3] to generate the training data. Blender can be controlled by the Python script, and Python 3.5.3 is shipped with the above-mentioned Blender. So, after we place the necessary objects and lights in Blender, the arbitrary number of training data can be generated automatically by using Python program according to the processes shown in Section 3. We created Blender objects and Python programs separately and executed them with the batch file shown in Fig. 7.

(a) of Fig. 8 shows the placement of objects in Blender, which is composed of a bowl, four marbles, and a square tray. In this experiment, same as the previous study, we used four types of marbles. So, we placed each one under the tray. And, the tray is used to detect the spillage of marble objects from the bowl object. That is since there is a possibility that the marble objects spill out in the physical simulation in the case where the number of them is large, such an image must be excluded. (b) of Fig. 8 shows the view from the camera for the rendering. The white solid quadrangle is the outline of the tray, and the lower right is the lowest position. So, the spilled marble objects gather here. Then, the spillage can be detected by comparing the image before and after the simulation after trimming the bowl object. Incidentally, as for the illumination, we placed Hemi lamp, which lights up the whole area equally, and a directional lamp on the side of the camera; the range between white dashed lines in (b) of Fig. 8 was the rendering area.

The procedure to generate the images for the training data by using the batch file in Fig. 7 was as follows. Firstly, the Blender file was opened and the marble objects in (a) of Fig. 8 were randomly selected, then their copies were placed hierarchically as shown in (a) of Fig. 6. Here, to shorten the falling time, each hierarchy was divided into four quadrants, and marble objects were placed randomly within the range of each quadrant. Next, the state of (b) in Fig. 6 was generated by physical simulation. Then, rendering was performed by the camera placed right above the bowl object as shown in (b) of Fig. 8, and its image was saved in a file. After that, the Blender file was reopened to return to the initial state, then the same procedure was repeated.

Table 1: Experiment cases on marbles.

No	Case	Position	Illumination	Material
(1)	Rough			
(2)	Position	◯		
(3)	Illumination	◯	◯	
(4)	Material	◯		◯
(5)	All	◯	◯	◯



(0) Real (1) Rough (2) Position

(3) Illumination (4) Material (5) All

Figure 9: Images of marbles by photography and CG.

After all the image data were generated, they are converted to the training data by another Python program by the following procedures. Firstly, images with spilled marble objects were excluded, then the outside of the bowl object in the remained images was trimmed to create images similar to Fig. 3. Then, the designated number of images were saved as the training data. After that, by the same procedure as the previous study as shown in Section 2, the images were padded 4 times to create the final training data.

5 EXPERIMENTS AND EVALUATIONS

In the experiments, firstly we evaluate each factor of creating CG images from the viewpoint of the influence on the accuracy of image recognition, by using marbles. Secondly, we evaluate the change of this accuracy according to the ratio of replacing a part of the CG images with the actual images. Thirdly, we evaluate the above-mentioned tendency for the object with different shape and material, by using nuts. In the above experiments, we use the CG images created in the procedure shown in section 4. Lastly, we evaluate the man-hours to create a part CG image only targetting the shape processing.

5.1 Evaluations of CG Images for Marbles

Table 1 shows the factors of creating the CG image, which was evaluated in the experiments: '(2) position', '(3) illu-

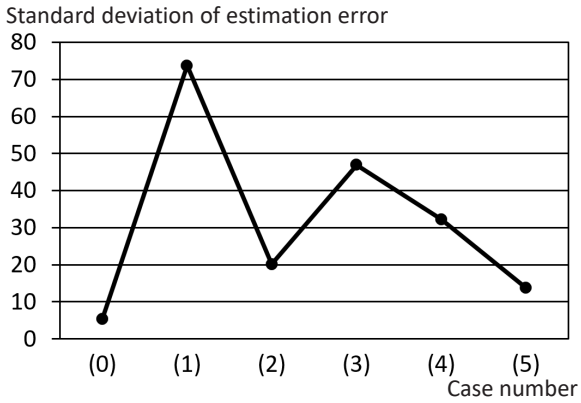


Figure 10: Change in standard deviation of errors.

mination' and '(4) material'. We evaluated the influence on the accuracy of each factor, and the case of combining them which is shown in '(5) all'. Incidentally, in Table 1, '(1) rough' is the state before improvement of each factor. Figure 9 shows the examples of the images of the training data of 30 marbles: '(0) real' shows the actual image and the others show the CG images created with each case of Table 1. Since marble objects were randomly placed in Blender in each case, the placements of the marbles were not the same.

In '(1) rough' of Fig. 9, since bowl modeling accuracy was low, the placement of marbles expanded and there were more gaps between marbles than the one in '(0) real'. So, in '(2) position', we improved the shape of the bowl, and the placement of marbles became closer to (0). Next, since the actual images were taken in a room where multiple fluorescent lamps were installed on the ceiling, we placed multiple lamps in Blender to make the CG images close to the actual environment in '(3) illumination'. In '(4) material', we improved only the material of the marbles from (2). In Blender, since it was necessary to change the rendering engine from 'Blender rendering' to 'Cycles rendering' in order to produce the marble material, the brightness of the whole image changed in (4). In '(5) all', we added the lamps of (3) to the CG image of (4).

Next, we trained the deep CNN shown in Fig. 4 with these training data. And, we prepared the test data composed of 50 actual images for each case of 5, 20, 40, 60 and 80 marbles. Then, we obtained the estimated quantity of each test data by the deep CNN. Figure 10 shows the standard deviation of the estimation error of each case of Table 1. This error is the difference between the actual quantity in each image and the estimated quantity using the deep CNN. Here, 'case number' corresponds to 'No.' in the Table 1. Although the standard deviation of '(0) real' was 5.3, the one of '(1) rough' worsened to 73.7. And, it was bettered to 20.1 by improving the position as shown in (2). On the other hand, each improvement of '(3) illumination' and '(4) material' worsened than (2). Here, each improvement was applied separately. However, in the case of applying both shown in '(5) all', it bettered to 13.8. That is, it was necessary to improve the illumination and material together.

Also, Fig. 11 shows the distribution of errors of estimated quantities with respect to each actual number of marbles in each case of between '(0) real' and '(5) all'. Here, since we

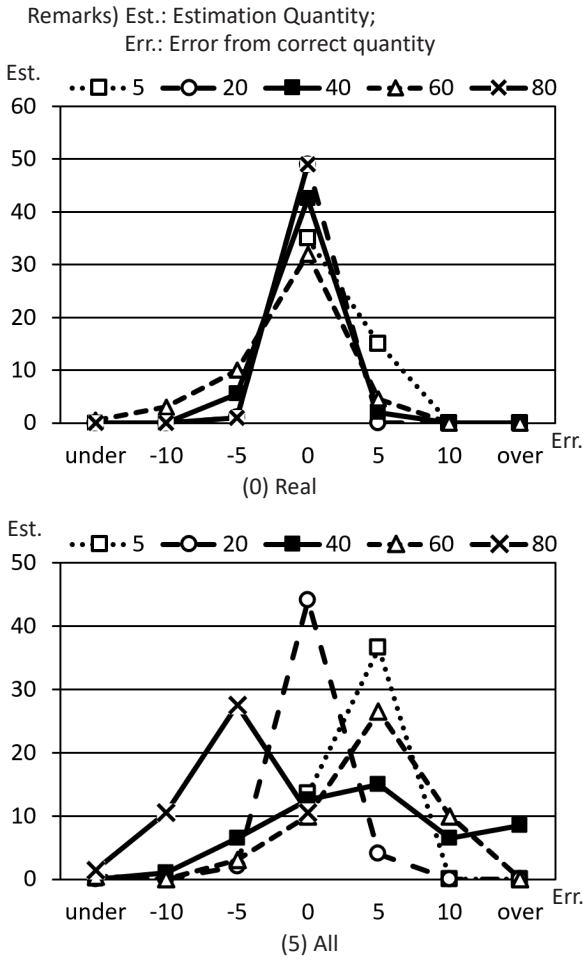


Figure 11: Distribution of errors of estimated quantities.

classified the number of marbles from 5 to 80 every 5 and trained by the supervised learning, the error also changed in units of 5. In (0), the error is distributed in the range of ± 10 around the correct '0'; while in (5), many peaks of distribution appeared before and after '0'. However, even in the latter case, though the error 'over' namely 15 or more occurred for about 20% images in the case of 40 marbles, the error was within the range of roughly ± 10 in the other cases.

5.2 Evaluations of CG and Actual Mixed Images

The creation of high-precision CG image similar to the actual images requires a large number of man-hours. Conversely, the actual images can be obtained even in the case of the above-mentioned inventory shelves in the factory, if the number of images is small. So, we evaluated the change of the accuracy in the case where the CG images are replaced with the actual images.

We used CG images of '(5) all' in Table 1 in this experiment, and the designated number of images were replaced with the actual images. For example, in the case of '5%', we prepared 500 images using 425 CG images and 75 actual images. In addition, the test data for evaluation are the actual images same as Section 5.1.

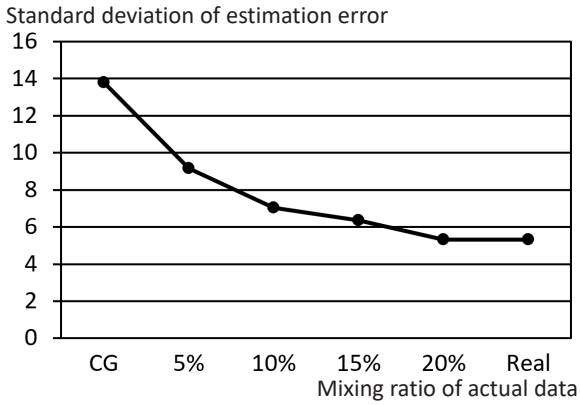


Figure 12: Standard deviation of errors on mixing ratio.

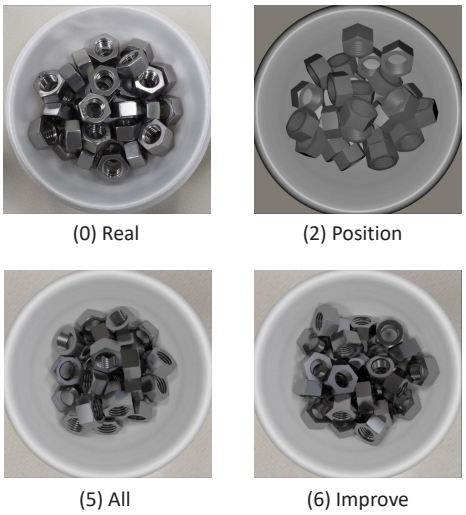


Figure 13: Images of nuts by photography and CG.

Figure 12 shows the change of the standard deviation of errors according to the mixed ratio of the actual data from 5% to 20% for every 5%. The standard deviation improved to about 2/3 when 5% of the CG images were replaced with the actual images, which was about 55% improvement as for the difference between (5) and (1). Similarly, as of 10%, it improved to about half, which was about 80% for the difference between (5) and (1). Furthermore, as of 20%, the standard deviation became almost the same as of the actual images.

That is, in the case of replacing the CG images with the actual images, the improvement of the accuracy was relatively larger than the replacement ratio.

5.3 Evaluations in Nuts

In order to confirm that even for the objects with different materials and shapes, they have the trends similar to those shown in Sections 5.1 and 5.2, we conducted the same experiment as that shown in Figs. 10 and 12. As shown in Fig. 3, the nut has a difference in material as well as shape from marbles, that is, it is made of metal.

Figure 13 shows examples of CG image of 30 nuts created for this experiment. Incidentally, '(0) real' is an example of actual image same as that in Fig. 9. Firstly, based on the results in Sections 5.1, we created '(2) position' that does not take material into consideration and has rough shape. This corresponds to (2) in Table 1, though the illumination is set similarly to '(5) all' in Table 1. Then, we created '(5) all' with a comprehensive improvement of material, lighting, and shape, similar to '(5) all' in Table 1. Lastly, since the shape of the nut is more complicated than the marble, we improved the shape again and created '(6) improve'.

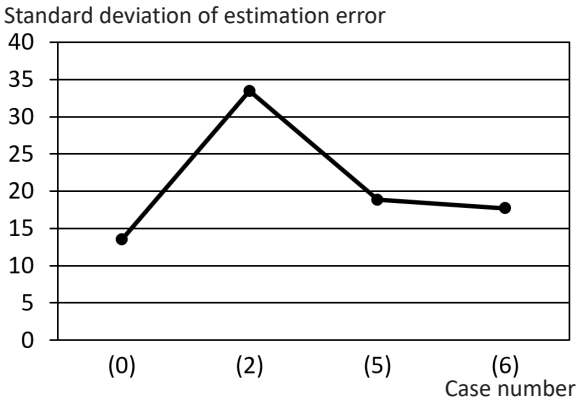


Figure 14: Change in standard deviation of errors (nuts).

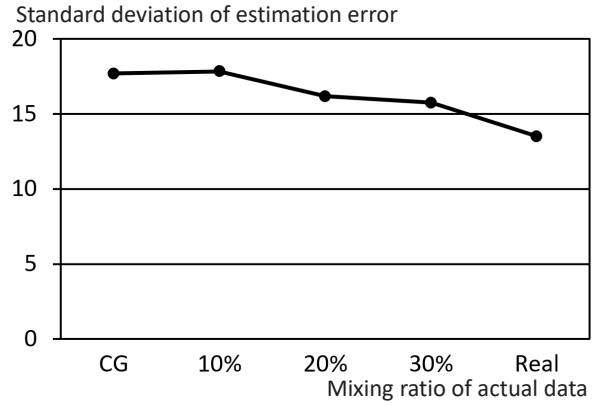


Figure 15: Standard deviation of errors on mixing ratio (nuts).

We trained the deep CNN with these data. Then, similar to Section 5.1, we prepared the test data with actual images of nuts and obtained the estimated quantity for each of them. Figure 14 shows the standard deviation of the estimation error, and we obtained the same tendency as that in Fig. 10. That is, it was 13.5 in the case of (0); it worsened to 33.5 in (2), and improved to 18.8 and 17.7 in (5) and (6) respectively.

Next, we evaluated the change of the accuracy in the case where the CG images are replaced with the actual images, similar to Fig. 12. As shown in Fig. 15, similar to the case of marbles in Fig. 12, accuracy was improved by replacing the CG images with actual images. However, its improvement rate was smaller than that of marbles.

5.4 Evaluations of Creation Efficiency of Part CG

Since factories handle a large number of parts, it is necessary to prepare these training data with CG images. There-

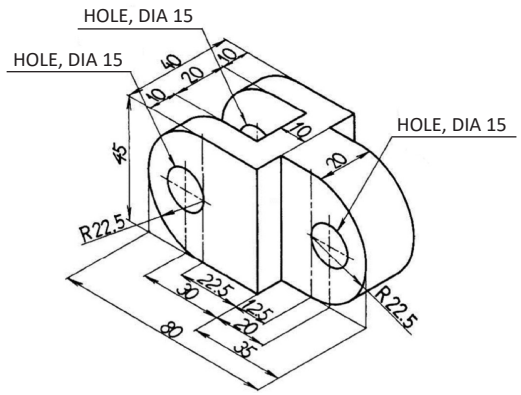


Figure 16: Drawing of machining. [11]

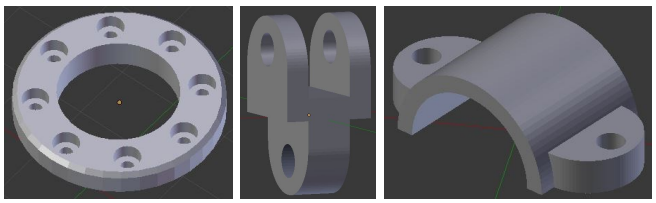


Figure 17: Parts modeling with CG tool.

fore, their quantity to be created also becomes large. On the other hand, from the viewpoint of CG image, there are few kinds of materials as shown in Fig. 5, and surrounding environment such as illumination is roughly the same. That is, as for one material, if there is a CG model of one part, it is only necessary to model the shape for the other part. Then, its training data can be created automatically by the procedure shown in Section 3. And, for the parts, there are drawing of machining shown in Fig. 16.

Therefore, on the premise of the same material and illumination as in Section 5.2, we evaluated the working time of parts modeling with the CG tool based on the drawing shown in Fig. 16. Figure 17 shows the CG model created in this experiment. Incidentally, '(2) part 2' is based on Fig. 16. Here, since target parts shapes are generally simple as shown in Figs. 1 and 5, we used drawings of a rudimentary machining [11].

Figure 18 shows the working time to create these models. Here, although the worker had some experience using

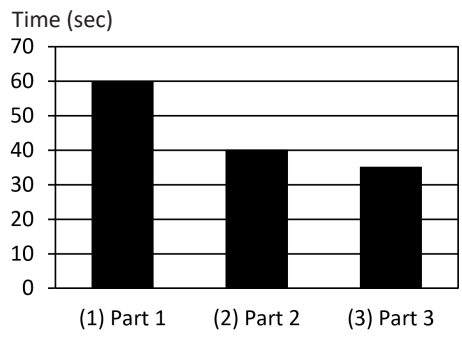


Figure 18: Working time for part modeling.

the CG tool Blender, he had no experience of parts modeling other than the above-mentioned nut, nor mechanical drawing. Therefore, the working time does not include the time used for interpretation of the drawing or consideration of the creation method. As shown in '(1) part 1' of Figs. 17 and 18, although the time increased as the processing place increased, the working time was 60 minutes at most.

That is, the working time to create the CG model was so smaller than that to take the picture mentioned in Section 2.

6 DISCUSSION

In this study, we evaluated the accuracy in the case of using the CG images instead of the actual images for the training data of the deep learning. Firstly, we examined the case of the marbles. As a result, as shown in Fig. 11, we found that a certain degree of accuracy could be achieved, while the errors spread larger compared to the case of the actual images. On the other hand, as shown in the relation between the CG images of Fig. 9 and the standard deviations of the error in Fig. 10, the accuracy was greatly affected by the factors to create the CG images.

In other words, it is expected that the accuracy can be improved by refining the CG image shown in (5) of Fig. 9 to make it closer to the actual images shown in (0). However, the more we refine the CG images to look like the actual images, the more the workload becomes big. That is, it is necessary to repeat the several works accompanied by trial and error: the first is the adjustments of the CG images such as material and illumination; the second is the deep learning of the deep CNN. And since each takes a long time, the ratio of improvement to cost is expected to decline.

Therefore, we showed that the accuracy could be improved efficiently by using the training data, in which a part of the CG images was replaced with the actual images, as shown in Fig. 12. That is, we found that in the case where the training data was generated with the CG images are used, it is effective to prepare the actual images in possible and to use them as a part of the training data. In addition, at the factory like the target of this study, the operation to estimate the inventory quantities by using the deep learning as follows is possible: initially, the training data is composed of the CG images and a few actual images; then, while increasing the actual images in operation, the training of the deep CNN is repeated by using these actual images and the accuracy can be gradually improved.

Here, to improve efficiently the accuracy obtained by using CG images, it is considered to use plural methods in combination: the proposed method in this paper, and another method, for example, a method using GAN shown in Section 2. By using the proposed method, the CG models of parts could be created accurately and efficiently from the drawings of machining as shown in Figs. 17 and 18; and, the placement of parts could be automatically determined by the physical simulation shown in Fig. 6. Therefore, since training like GAN is not necessary, we consider that this method is more efficient than GAN as for these processes. On the other hand, for example, to adjust the materials and illuminations of parts shown in Figs. 9 and 10, we had to perform trial and error.

So, we consider that the accuracy may be improved more efficiently by applying such as GAN to these processes, and this is one of the themes of the next study.

Furthermore, we examined the case of the nuts and obtained a similar result to the case of marbles. So, we consider that it is possible to prepare the training data for the deep learning by using CG images for parts of various materials and shapes. However, as shown in Figs. 12 and 15, we found that there was the following difference between the case of marbles and nuts. One was the value of the standard deviation of the error; Other one was the improvement ratio in the case of replacing the CG images with the actual images. And with regard to the cause of the latter, we considered that it was due to the ratio of errors between the case of using CG images and actual images, which was smaller in nuts compared to marbles.

Lastly, we evaluated the working time of CG image creation, we confirmed that the training data could be created far more efficiently than collecting actual images in the case of target parts. Therefore, it is considered that even the fields where it is difficult to accumulate a large number of actual images for the deep learning, there are fields where the target is composed of the simple objects from the viewpoint of CG. In such a case, we consider that to use the CG images generated automatically for the training data is effective.

However, in this study, we have verified the effectiveness of the proposed method by using only the marbles and nuts at just the laboratory. So, the verification in the actual factory remains as the future study. That is, it is necessary to evaluate the accuracy in the case of utilizing the CG images of the various shapes of parts in the actual inventory shelf environments such as shown in Fig. 1.

7 CONCLUSION

Currently, the accuracy of image recognition utilizing the deep learning is rapidly improving, and such an image recognition is applied to various fields. On the other hand, in order to improve the accuracy by the deep learning, it is necessary to accumulate a large number of training data. And, the preparation of the training data often becomes the obstacle to applying the deep learning.

For this problem, we proposed a training data generation method by utilizing CG in this study. And, we created the CG model and confirmed that a large number of training data could be generated by using the CG tool automatically; a certain degree of the accuracy could be achieved by using such a training data. Also, we found the accuracy could be improved by replacing a part of these training data with the actual images. Furthermore, we confirmed training data could be created much more efficiently by this method than collecting the actual images in the case of simple objects such as mechanical parts. As a result, we conclude it is effective to generate the training data by using the CG images in some fields, where it is difficult to accumulate a large number of the actual images for the deep learning.

Future study will focus on the confirmation of its effectiveness in actual inventory environment, namely by using the actual parts in the actual factory, and the method to improve the accuracy efficiently, especially due to the material and illumination.

REFERENCES

[1] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," Computer networks, Vol. 54, No. 15, pp.2787–2805 (2010).

[2] K. Banker, "MongoDB in Action," Manning Pubns Co. (2011).

[3] Blender Foundation, "Blender," https://www.blender.org/ (reffered, May 2018).

[4] T. H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "PCANet: A simple deep learning baseline for image classification?," IEEE Transactions on Image Processing, Vol. 24, No. 12, pp. 5017–5032 (2015).

[5] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," Mobile networks and applications, Vol. 19, No. 2, pp.171–209 (2014).

[6] F. Chollet, "Deep learning with python," Manning Publications Co. (2017).

[7] D. Ciregan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," 2012 IEEE conference on computer vision and pattern recognition (CVPR), pp. 3642–3649 (2012).

[8] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," IEEE Transactions on Medical Imaging, Vol. 35, No. 5, pp. 1153–1159 (2016).

[9] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, "Deep learning," MIT press (2016).

[10] I. Goodfellow, "NIPS 2016 tutorial: Generative adversarial networks," arXiv preprint arXiv:1701.00160 (2016).

[11] M. Higuchi, "Drafting basics," http://www3.nit.ac.jp/ mhiguchi/seizu-01.pdf (referred Oct. 10, 2018) (in Japanese).

[12] T. Kawanaka, and T. Kudo, "Inventory Satisfaction Discrimination Method Utilizing Images and Deep Learning," Procedia Computer Science, Vol. 126, 937-946 (2018).

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton,"Imagenet classification with deep convolutional neural networks," Proc. advances in neural information processing systems, pp. 1097–1105 (2012).

[14] T. Kudo, Y. Ito, and Y. Serizawa, "An Application of MongoDB to Enterprise System Manipulating Enormous Data," Int. J. Informatics Society, Vol. 9, No. 3, pp. 97–108 (2017).

[15] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," Frontiers in plant science, Vol. 7, Art. 1419 (2016).

[16] E. Redmond, and J.R. Wilson, "Seven Databases in Seven Weeks: A guide to Modern Databases and the NoSQL Movement," Pragmatic Bookshelf (2012).

[17] F. Seide, G. Li, and D. Yu, "Conversational speech transcription using context-dependent deep neural networks," Proc. twelfth annual conference of the interna-

tional speech communication association, pp. 437–440 (2011).
[18] E. A. Silver, D. F. Pyke, and R. Peterson, "Inventory management and production planning and scheduling," John Wiley & Sons (1998).
[19] S. P. Singh, "Production and Operation Management," Vikas Publishing House Pvt Ltd (2014).
[20] M. Verhelst, and B. Moons, "Embedded Deep Neural Network Processing: Algorithmic and Processor Techniques Bring Deep Learning to IoT and Edge Devices," IEEE Solid-State Circuits Magazine, Vol. 9, No. 4, pp. 55–65 (2017).

(Received October 10, 2018)
(Revised February 11, 2019)

Tsukasa Kudo received the BSc and ME from Hokkaido University in 1978 and 1980, and the Dr. Eng. from Shizuoka University in 2008. In 1980, he joined Mitsubishi Electric Corp. He was a researcher of parallel computer architecture and engineer of business information systems. Since 2010, he is a professor of Shizuoka Institute of Science and Technology. Now, his research interests include deep learning and database application. He is a member of IEIEC and IPSJ.

Ren Takimoto is currently working toward a B.I. degree at Shizuoka Institute of Science and Technology, Japan. His research interests include computer graphics design.

Tenma Kawanaka received a B.I. from Shizuoka Institute of Science and Technology, Japan, and joined Shizutetsu Information Center Corporation in 2018. His research interests include deep learning. His contribution to this paper was carried out in his graduation research.

Industrial Paper

A Study on Time Synchronization Method for Field Servers for Rice Cultivation

Koichi Tanaka†, Mikiko Sode‡, Tomochika Ozaki†, Masakatsu Nishigaki†, Tadanori Mizuno*

†Graduate School of Science and Technology, Shizuoka University, Japan
‡Global Information and Management, International College of Technology, Japan
*Faculty of Information Science, Aichi Institute of Technology, Japan

Abstract - It is important to develop affordably priced field servers for rice farmers for their practical implementation. This restricts the use of expensive GPS or high precision crystals, which have been used for wireless communication so far. To this end, we propose a time synchronization method that does not involve the use of expensive hardware. In the field servers for rice fields that use LPWA technology, which require only batteries for their operation, time synchronization is an important factor in reducing power consumption. Therefore, we describe a method of constructing a wireless network of an economical time-synchronized field server using LoRa for achieving low costs. We also describe the effect of reducing power consumption. From experimental results, we confirmed that the time was synchronized and transmission and reception of data between the master unit and the field server ensued normally. It is theoretically possible to operate the device for 691 consecutive days. In addition, we confirmed that the field server system works correctly from rice planting to rice reaping in the rice field.

Keywords: Agriculture, Field Server, Sensor Network, Low Battery Consumption, Time Synchronization

1 INTRODUCTION

The average age of farmers is rising in Japan [1]. Therefore, reduction of labor burden is an important issue. To reduce labor burden, the introduction of a field server that can reduce the man-hours required for daily field surveillance could be effective. Field servers are available in the market. However, their use is not prevalent owing to the associated high cost. Thus, affordable field servers that rice farmers can purchase are urgently required.

We are developing a field management system that can help reduce labor burden [2]. To introduce the field server to farmers, it is important to reduce the installation and operation costs. Therefore, we adopted LoRa, for its advantages of low power consumption and long communication distance, which make it suitable for communication in the field server for the rice fields [3]. In addition, LoRa does not incur any communication cost.

To reduce installation costs, it is important for the device to be operable using batteries because it is expensive to draw power to rice fields. In addition, it is difficult to install solar panels because solar panels are large and may interfere with farm work. The field server must be able to be battery-operated for at least six months, starting from rice plantation to reaping. Therefore, low power consumption is important for the field servers for rice fields.

To operate for six months with no power supply, the field server needs to turn itself off except when sending or receiving the sensor data or other communication. This requires an intermittent operation communication protocol and a time synchronization method to be implemented.

The time synchronization technique has been extensively studied previously [4, 5]. The method for time synchronization using GPS has also been proposed [4]. To use this method, it is necessary to install a GPS receiving module in every field server, which leads to an increase in the initial cost at the time of installation. The power consumption also increases and, therefore, such a method is not suitable for use in the rice field servers, for which lowering the introduction barrier is desired. Although TPSN [5] has been proposed as the time synchronization method, it requires a long time for time synchronization, which increase the power consumption, and this, this method is not suitable for use in field servers.

To realize low power consumption, intermittent operation is indispensable. Time synchronization is an important technique to ensure that a plurality of field servers operate efficiently by employing intermittent operation. To communicate efficiently, a communication partner must be starting up. In this study, we propose an intermittent operation communication protocol and a time synchronization method to solve the aforementioned problems. In the proposed method, after the field server system transmits the sensor data to the master unit system, the master unit system, on receiving the sensor data, transmits the time correction signal to the field server system, thereby performing time synchronization.

To realize affordable price, it is not possible to use expensive GPS or high precision crystals, which have been used for wireless communication so far. Therefore, we propose a time synchronization method that does not use expensive GPS and high precision crystals. The proposed method has a mechanism to allow time variation of the field server and use it for improving the reliability of the system. In the proposed method, a large frame is taken to allow variations, and retransmission processing can be performed using the margin.

In this paper, section 2 discusses the necessity of affordable price field server, section 3 describes the limitations of the conventional time synchronization systems, section 4 describes the system configuration, section 5 explains the proposed communication protocol and time synchronization method, section 6 presents the operational test and result, and finally, section 7 summarizes the study.

2 NECESSITY OF AFFORDABLE PRICED FIELD SERVER

In Japan, aging of farmers has progressed; the average age of the agricultural working population was 66.8 years in 2014 [1]. In addition, agricultural employment population is decreasing. Currently, agriculture in Japan is in a crisis situation, and it is necessary to increase the number of agricultural workers, for which it is necessary to reduce the burden on the workers. Therefore, acquiring and using the environmental data is considered important. However, field server introduction is not progressing.

In Japan, rice farming generates less revenue compared to other crops. Table 1 presents the profit structure in the rice production revenue. Tan (反) and pyou (俵) are Japanese units of measurement. 1 pyou equals 60 kg, and 1 tan equals 0.1 ha. The standard amount of rice harvested per tan is 9 pyou. One pyou rice can be sold for approximately 13,000 yen. Deducting expenses will result in a profit of approximately 50,000 yen per tan. On the other hand, a field server can be hired at 8,280 yen per month [6]. If the server is rented for 6 months, the cost becomes 50,000 yen, which implies that almost no profit is obtained. Therefore, the introduction of field servers is difficult.

To facilitate practical implementation of the field server, it is necessary for its selling price to be less than 10,000 yen. In other words, it is necessary for the manufacturing price to be approximately 3,000 yen. Therefore, we propose a time synchronization method for field servers, which can help realize cost reduction. We aim to develop protocols for affordable and manufacturable field servers.

The field server should not interfere with farm work. Because large agricultural machines operate in rice fields, the field server needs to be moveable. Therefore, the height of the server should ideally be 1 m or less and it needs to be as compact as a lunch box. Also, because a rice field has no power supply, the device needs to be operable using batteries, from the phases of rice planting to harvesting.

To realize an affordably priced field server, a communication line usage fee is unnecessary, and a LoRa network, which can be transmitted to large distances, is used. LoRa is more effective for IoT in agriculture, as a larger amount of data can be transmitted compared to Sigfox [7], and a relatively large amount of sensor data is involved. Further, it is suitable for special applications in which performance and cost are critical factors, because we can freely create protocols and frame formats. LoRa can be built and operated on its own, including base stations, and its specifications are open.

Table 1 Profit structure in rice production.

60kg(1 俵)	¥13, 000-
Rice crop yield/1 反	9 俵
Revenue/1 反 (Fertilizer ¥15,000- / 1反 Herbicide / pesticide ¥10,000- / 1反 Land improvement expenses for canal improvement and irrigation etc. ¥20,000- / 1反 Labor cost ¥15,000- / 1反)	¥57,000- (Expense ¥60,000-)

To create a field server for use in rice cultivation, it is necessary to reduce the number of high-cost parts. Expensive crystals and GPS cannot be used for time synchronization because of the associated high cost. It is thus necessary to realize time synchronization without these parts.

The operational target is to realize server operation in the rice field of Ishikawa Prefecture's second largest agricultural corporation. The field servers acquire the sensor data once every hour and send it to the cloud.

3 PROBLEMS OF CONVENTIONAL TIME SYNCHRONIZATION SYSTEM

The use of wireless smart utility networks (Wi-SUN) [8] was standardized as the wireless communication method for metering electricity, gas, and water, around the year 2008. Before this, other standardized wireless communication specifications such as ZigBee were proposed to realize a sensor network [9]. The difference between these and Wi - SUN is that Wi-SUN attempts to enable wireless communication in a wide area. Figure 1 shows the MAC protocol architecture of IEEE 802.15.4 / 4e. The MAC protocol of IEEE 802.15.4 / 4e, can roughly be divided into asynchronous and synchronous networks. Synchronous networks have better power efficiency and lower power consumption than asynchronous networks. That is, in IoT, in which low power consumption is essential, the synchronous system is more suitable than the asynchronous system.

The synchronization method can be roughly categorized as the beacon method and channel hopping method. Figure 2 shows an overview of the IEEE 802.15.4 beacon superframe method. This method is basically a better way when all nodes can receive signals during the beacon period. The field server we propose performs intermittent operation to minimize power consumption. When the power supply of the field server is turned on at an arbitrary time on the premise of intermittent operation, it is necessary to prepare a standby time for receiving the beacon separately from the transmission operation of the measured data, and the extra power is consumed. Therefore, it is difficult to use this method for time

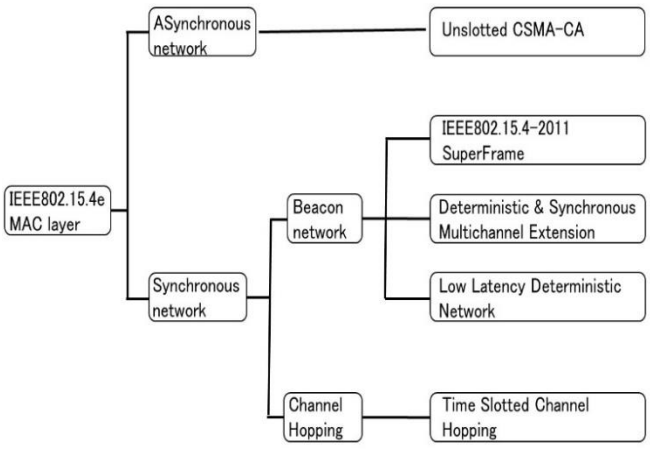


Figure 1: MAC protocol classification of IEEE 802.15.4 / 4e. [8]

synchronization of our system. Figure 3 shows the outline of the IEEE 802.15.4 channel hopping TSCH method. In the beacon superframe method, the beginning of a superframe is determined by a beacon (synchronization control signal), and all the terminals receive it. All terminals then synchronize with each other. On the other hand, in channel hopping, TSCH shares the time with each node with accuracy within tens of microseconds in the whole network. Each node maintains synchronization by exchanging information on the timing error with adjacent nodes. These methods require highly accurate time synchronization, and require GPS and highly accurate crystals, etc., which increase the cost substantially. This standard is therefore unsuitable for use as the communication standard for rice cultivation.

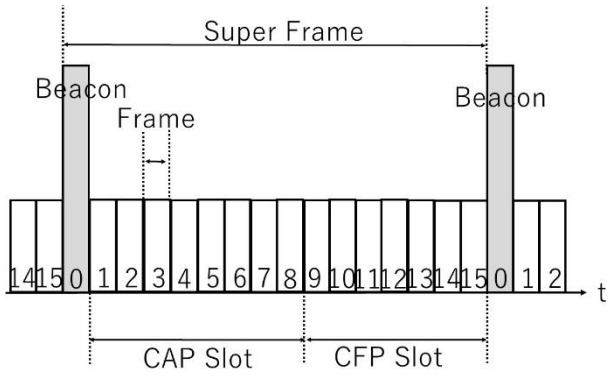


Figure 2: Outline of IEEE 802.15.4 beacon · superframe method.

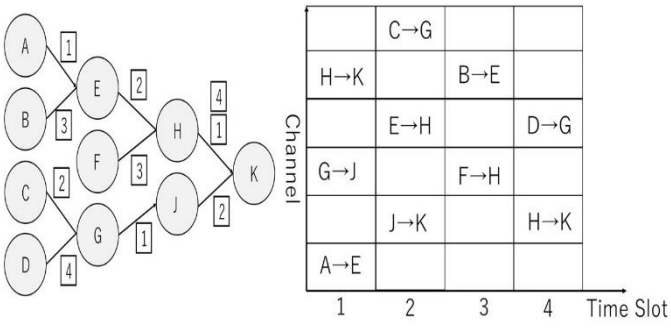


Figure 3: Outline of IEEE 802.15.4 channel hopping · TSCH method.

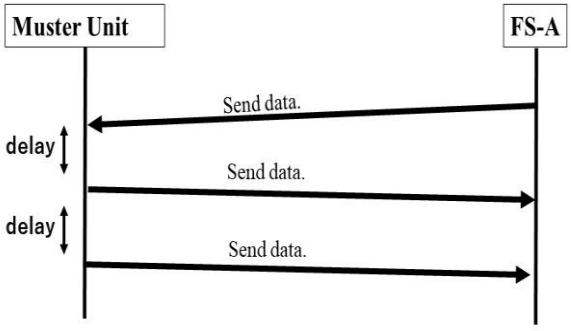


Figure 4: Sequence of LoRaWAN Class A.

The LoRaWAN has three device classes. Class A can transmit data at an arbitrary timing; reception is possible only immediately after transmission, and it is a power saving device that goes to sleep immediately after reception is completed (Figure 4). Since the data is transmitted at an arbitrary time, there is a possibility of collision. In addition to the function of the class A device, the class B device has an arbitrary timing reception mode, and it is a device that can receive the beacon from the master unit; however, the power consumption becomes higher than that of the class A device. Class C is a device that can constantly receive data, and its power consumption is exceedingly high. The time synchronization method of LoRaWAN is basically the same as that of the Wi - SUN. Therefore, even in LoRaWAN, GPS or high accuracy crystals are needed.

4 SYSTEM CONFIGURATION OF FIELD MANAGEMENT SYSTEM

The field management system is composed of the field server system, master unit system, and cloud service. Figure 5 shows the overall structure of the field management system. The field server system is installed in the rice fields and receives the sensor data. Further, the data is sent to the master unit system through the LoRa wireless network. The master unit system integrates the sensor data from the field server system and sends them to the cloud service through the 3G line or Wi-Fi. The cloud services are provided by the smartphone applications, tablet applications, and web pages. These services provide data to farmers to alert them about water levels, propose a suitable work plan, preserve work records etc.

Communication between the field server system and the master unit system using LoRa is capable of long-distance communication. LoRa has been found to have a practical communication distance of 3,000 to 4,000 m as shown by the basic communication characteristics survey [10]. The rice field of Ishikawa prefecture was assumed, and the linear distance between the master unit and the field server was within 3,000 m. For this reason, we adopted LoRa, which enables direct communication between the field server and master unit.

Figure 6 shows the positional relationship between each field and the office of the assumed the agricultural corporation in the Ishikawa prefecture. A, B, C, D, E, F, and G represent the position in each field, where the field servers are installed. P represents the location of the office; the master unit is installed in the office. The linear distances between the field servers A to G and the master unit P are as follows: 397 m between A and P, 923 m between B and P, 943 m between C and P, 684

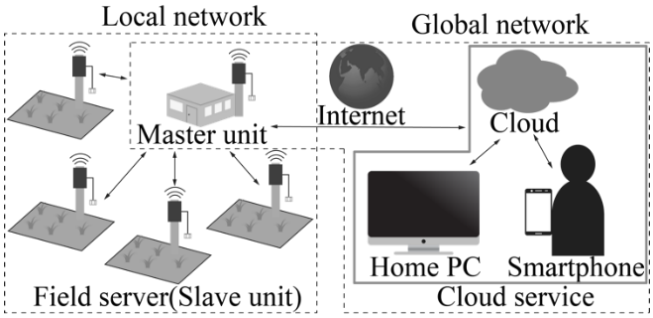


Figure 5: Overall structure of rice cultivation management system.

m between D and P, 1,150 m between E and P, 1,440 m between F and P, and 1,910 m between G and P.

The location and number of field servers installed were examined by the agricultural corporations. The decision method was set as a place to be checked whenever looking around done every day. If it was possible to confirm the water level etc. at the designated place, it was that it was enough for management of the field.

We will further explain the configuration of the field server system installed in the rice field and the master unit system installed in the office. Figure 7 shows the configuration of the field server system. The field server comprises the battery, power ON/OFF circuit, AVR microcomputer, LoRa module, various sensors, and SD card module. The field server is powered by the battery. To realize low power consumption, the power ON/OFF circuit operates only for several tens of seconds in one hour. The wireless modules and the sensors are controlled by the AVR microcomputer. Five types of sensors are mounted to measure the temperature, humidity, water level, soil temperature, and soil moisture content. The sensor data is stored in the SD card together with the time stamp. This is a function for reliably saving the data, considering the case where it cannot be transmitted to the master unit or where the time correction signal cannot be received. The power ON/OFF circuit is composed of the PIC microcomputer and the FET; it controls power supply to the AVR microcomputer. The PIC microcomputer controls the FET by outputting HIGH/LOW at the GPIO pin. The time required for the power supply control is calculated and controlled by using the timer interrupt in the internal clock of the PIC microcomputer.

The configuration of the master unit system is shown in Figure 8. The master unit system is composed of a Raspberry Pi, LoRa module for transmission, LoRa module for reception, and a 3G dongle. When the field server system is turned on

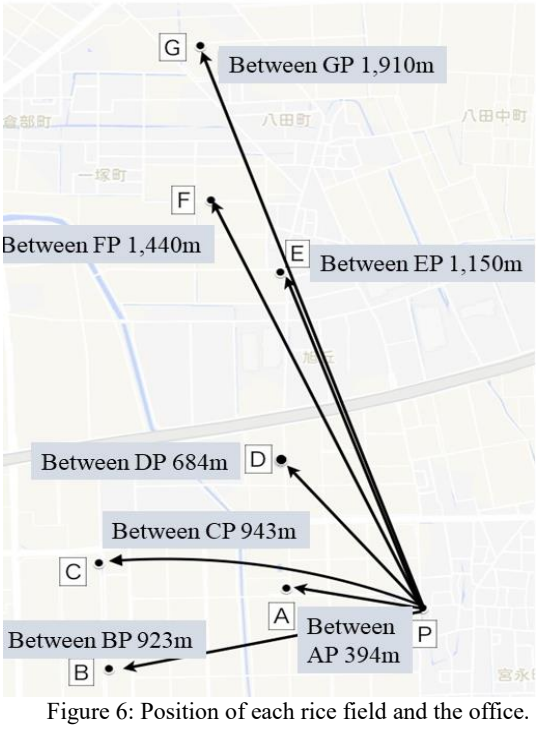


Figure 6: Position of each rice field and the office.

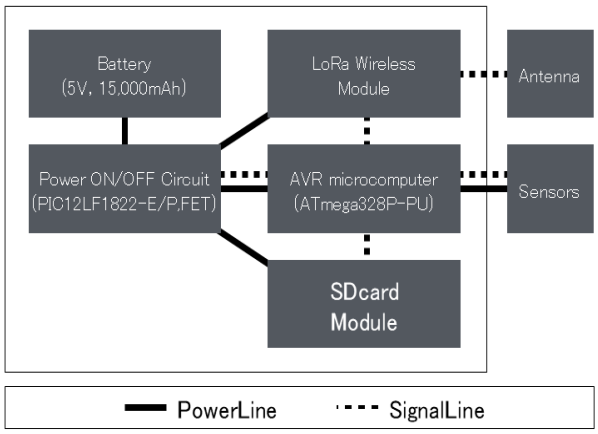


Figure 7: Field server system configuration.

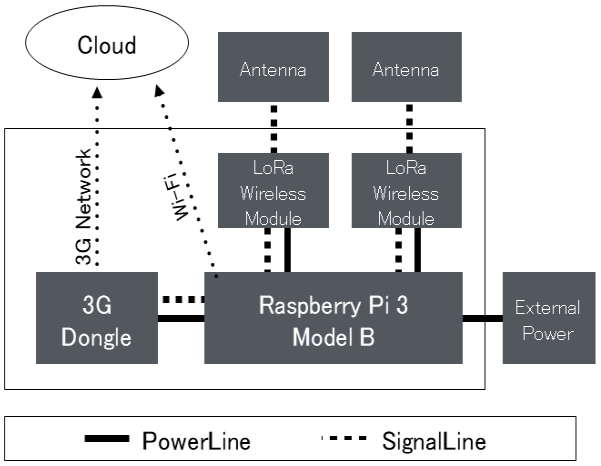


Figure 8: Master unit system configuration.

for the first time, the time is not held and the time is set after it is transmitted by the master unit system. Therefore, the master unit system always maintains the reception state. It is desirable that the master unit system can respond to communication from the field server system when the field server is installed. Further, the reception and transmission modes exist in the LoRa module, and it takes time to switch the modes. Therefore, by installing two different LoRa modules for receive and transmit, it is possible to reduce the waiting time of transmission and reception and maintain the reception state at all times.

5 COMMUNICATION PROTOCOL

5.1 Communication Protocol

The frame formats used for the communication are shown in Tables 2, 3, and 4. Table 2 shows the common frame format, consisting of the destination, the source, and the payload. Table 3 shows the format of sensor data transmission. Since there are five types of sensors in use, the sensor data are defined in the format of 1 to 5. The temperature, humidity, water level, soil temperature, and soil moisture content are entered in that order from the sensor data 1 to 5. The acquired five types of data can be stored in 2-byte units. Additionally, when

Table 2: Common frame format.

Destination	Source	Payload
1Byte	1Byte	Variable

Table 3: Format of sending sensor data (Payload).

Sensor Data 1	Sensor Data 2	Sensor Data 3	Sensor Data 4	Sensor Data 5
2 Byte	2 Byte	2 Byte	2 Byte	2 Byte

Table 4: Format of correction time signal (Payload).

Timestamp (UNIX Time)	Correction time
4 Byte	2 Byte

the number of types of sensors increases, 2 bytes are added to the format of sensor data transmission. Table 4 shows the format of the time correction signal. The time stamp and the time correction signal transmitted from the master unit to the field server are stored in the payload. The field server system starts once every hour. It gets the sensor data and transmits it to the master unit system. When transmission is successful, the master unit system sends the correction time to the field server system. The field server system further corrects its own internal clock for the time synchronization. Later, when the field server system receives the corrected time or go on operating time per hour of described later elapses, the power is turned off except for the power control circuit.

Sensor data is acquired at the same time in all filed servers. Therefore, the sending data arbitrarily will conflict with each other. To prevent conflicts, each field server sends the data to the master unit in order. As a result, since the data can be efficiently transmitted to the master unit, power consumption can be reduced. The series of actions shown in figure 9 is basically done within the frame. The frame of each node is made to be large so that it can retransmit several times. The part that controls the transmission order of each node is a big difference from LoRaWAN. Figure 10 shows an example of three field servers in which resending mode does not occur in any of the communications. First, the field server system A (hereinafter, FS-A) is activated. FS-A measures the sensor data and generates the sending packet according to the sensor data. Further, the packet is sent to the master unit system. The field server system has only one LoRa module, therefore it switches from the sending to the reception mode. This switching requires several seconds. After switching to the reception mode, the field server

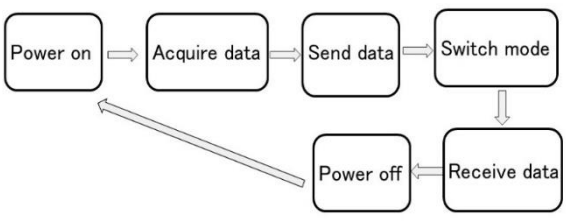


Figure 9: Communication protocol sequence.

system waits until the set timeout period. Further, it receives the corrected time signal from the master unit system. When the master unit system receives sensor data from FS-A, it sends the corrected time signal to FS-A. FS-A corrects its internal time based on the received correction time signal. After the correction, FS-A goes into sleep mode even during the resending possible time. When the field server system fails to receive the corrected time signal or the master unit system fails to send the corrected time signal to the field server system it is necessary to resend it along while the possible resending time. Field server system B (FS-B) and field server system C (FS-C) perform in the same sequence as A.

The operating time per hour can be obtained from equation (1).

$$\begin{aligned}\text{operating time} = {} & \text{sensor data acquisition time} + \\ & \text{sensor data transmission time} + \\ & \text{time correction signal reception time} + \\ & \text{retransmission time} \end{aligned} \quad (1)$$

This operating time is written in advance to the AVR microcomputer and sent to the PIC microcomputer each time the AVR microcomputer is powered on. The PIC microcomputer uses this value to calculate the restart time. Here, the sensor data acquisition time is the time to measure the sensor data. The sensor data transmission time is the time to transmit the sensor data to the master unit. The time correction signal reception time is the time to receive the current time from the master unit. The retransmission time is the time to perform retransmission processing when sensor data cannot be transmitted to the master unit. In the proposed method, time synchronization is performed once every hour. It has been confirmed from the measurement results that there is an error of more than ± 10 seconds at the maximum in an hour [11]. Therefore, the error of acquisition time of sensor data is also about ± 10 seconds. This error is a problem-free range as the sensor acquisition time error for agriculture.

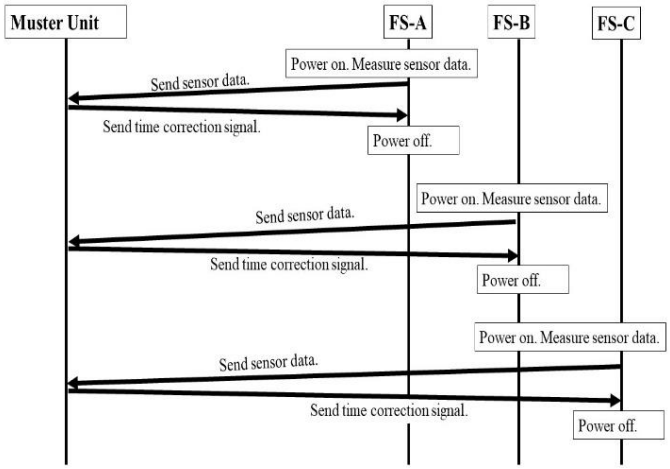


Figure 10: Communication protocol sequence.

5.2 Operation of Resending Mode

When the field server cannot receive the time correction signal and the reception waiting time has elapsed, the field server performs the timeout operation. After the timeout, the field server waits for random seconds from 0.1 to 5.0 s and then retransmits. The following is the cause of the timeout.

1) When the master unit cannot receive the communication from the field server due to the radio wave attenuation
2) When a collision occurs in the transmission data due to overlapping of the transmission times of a plurality of field servers
3) When the field server cannot receive the communication from the master unit due to the radio wave attenuation

The first one and third one are that the cause of the noise is often temporary, so there is a high possibility that the problem will be solved if the transmission process is performed with shifted time. The second one can be prevented by accurate time synchronization.

When the noise or the collision occurs, the master unit does not send the time correction signal to the field server, therefore, the field server's reception standby timeout occurs. The field server that has timed out executes retransmission, but to prevent re-collision with the communication performed by the field server of the initial power-on, a random second standby time is provided. After executing the retransmission, the field server switches from the transmission to the reception mode and waits for reception. This operation is continued until the field server can receive the time correction signal from the master unit. However, to avoid collision of the communication with that of other field servers, the power is turned off forcibly when the other field server's operation is about to start.

The sequence operation in this case is shown in figure 11. In figure 11, the field server C (FS-C) has timed out and is retransmitting. If it is within the possible retransmission time, the processing of the transmission and reception standby is repeated until transmission/reception is completed.

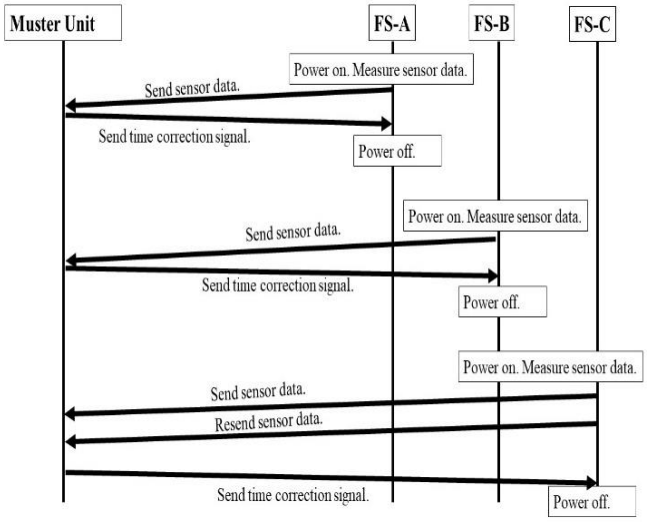


Figure11: Sequence for transmitting sensor data.

5.3 New field server installation mode

This section describes the operation procedure when a new field server is introduced. Figure 12 shows the process of installing a new field server. The newly added field server first sends a wake-up signal to inform the master unit that it has been added. When the master unit gets the wake-up signal from the field server, the field server number is added to the library and the current time, and next activation time is transmitted to the field server. The field server synchronizes the time based on the received time from the master unit.

The time to install the field server is arbitrary. Therefore, when another field server and the master unit are communicating, a newly added field server may start communication. In this case, collision occurs (see Figure 13).

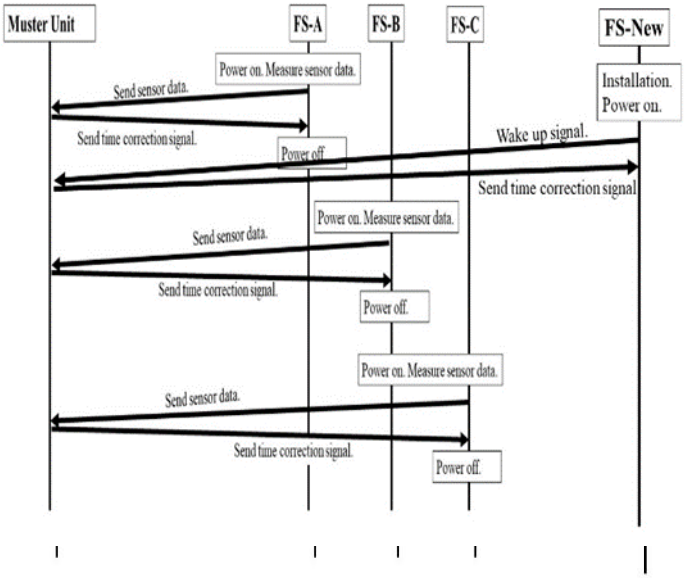


Figure 12: Sequence of new field server installation.

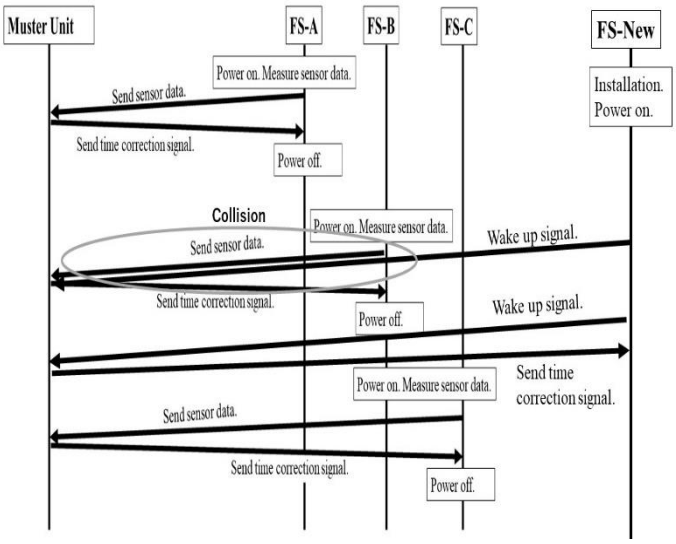


Figure 13: Sequence at collision in new field server installation mode.

When collision occurs, retransmission is performed after a random period within seconds. This process is repeated until the unit is receiving the current time information from the master unit. When the additional field server receives the current time information from the master unit, the time to acquire the sensor data (next time to turn on the power and acquire the sensor data) is calculated, and power is turned off.

5.4 Time Synchronize Signal

The field server can transmit the sensor data at an arbitrary timing because the master unit is always on standby for reception. However, in the case of the rice field management, since sensor data is acquired at the same time in all field servers and data is transmitted to the master unit, transmission from the field server to the master unit occurs at the same time, causing collision and the efficiency decreases. Therefore, in the protocol for the rice cultivation, the master unit performs the scheduling and notifies the time to transmit to the field server, and the field server basically transmits the specified time data. Figure 14 is an example of the scheduling. FS - A, FS - B, FS - C, FS - D are scheduled to send the data to the master unit in order. It is also possible to lengthen the allocation time for places where it is relatively difficult to transmit besides the building or sideways of the expressway and there is a high possibility of retransmitting several times.

The lower section in Figure 14 shows an example in which dispersion occurs with respect to the scheduling result from the above figure. Depending on the field server, the variation can be arbitrary, and there are several ranges and directions of variation. Therefore, although the probability is low, collisions may occur. In this example, FS - B shifted in the direction of lag, and FS - C shifted in the direction to become faster, and thus, overlapping occurred. Therefore, until FS - B processing is finished, FS - C must wait while sending it; however, because each frame has sufficient margin for variation, data can be sent without any problems.

Figure 15 is a diagram showing the state of retransmission when the collision of figure 14 occurs. FS - B is delayed by 12 seconds, FS - C and FS - D are 12 seconds earlier. The transmission periods should not overlap. Therefore, FS - C will retransmit after a random time of 0.1 to 5 seconds. In this example, retransmission occurred after 5 seconds. The transmission processing of FS - B has already been completed, so FS-C can be sent. The transmission of FS-C takes about 12 seconds, but since the transmission of FS-C is completed before FS - D starts transmission, no conflict occurs between FS - C and FS - D. Even if variations occur in such away as to interfere with the transmission and the reception in this case, it is a frame length setting that can be retransmitted sufficiently.

The master unit performs the time synchronization with the NTP server beforehand and acquires accurate time. When the master unit receives sensor data from the field server, the field server calculates the time to transmit next and informs the field server. Simultaneously, a time correction signal is also transmitted. The master unit informs the time to send the next data to the field server, that is, the scheduling result by this processing.

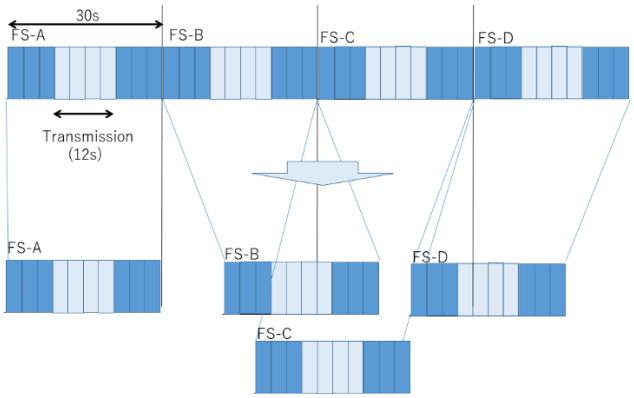


Figure 14: Scheduling and field server time variation.

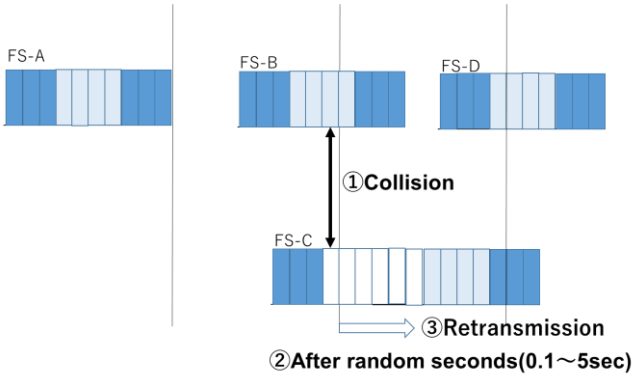


Figure 15: Retransmission processing by collision.

The format of time correction signal is shown in Table 4. The time stamp is entered with 4-byte UNIX time. It is used to write the sensor data to the EEPROM or the SD card of the AVR microcomputer. Since the correction time is used for correcting the time within the PIC microcomputer, the time shifted for each field server from the current time is set as 0 to 3599 s in 2 bytes. CRC etc. is used for detecting and correcting communication errors that are not defined in the format because they are added by the LoRa communication module.

Figure 16 shows the mechanism of the time synchronization between the field server and master unit. The master unit sends the correct time obtained by the NTP server to field server in the corrected time format. Upon receiving the correction time, the field server transfers the correction time via the AVR microcomputer to the PIC microcomputer in the power ON/OFF circuit. To prevent the correction time from starting simultaneously with other field servers, the current time is shifted appropriately according to the scheduling result.

In the example of Figure 15, the frame length is set to 30 seconds. It takes about 12 seconds for the master unit to receive data from the field server, change the communication mode, and send the time data to the field server. The frame length is set to 30 seconds by adding ± 9 seconds, taking into consideration the time variation of the PIC microcomputer.

In the case of the frame length is 30 seconds, a formula for calculating the start time is shown in equation (2). Regarding equation (2), each field server has a uniquely assigned field server identifier (below), and the correction time to transmit

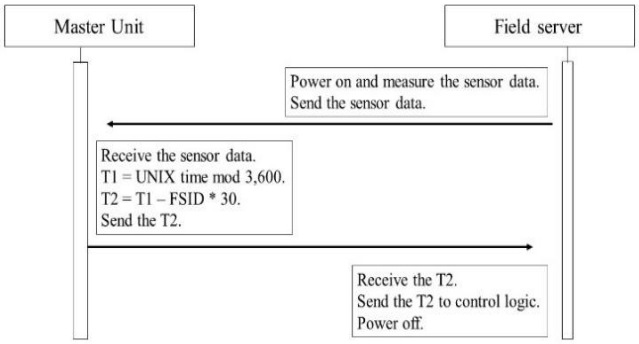


Figure 16: Time synchronization mechanism between master unit and field server.

to the field server is obtained by subtracting from the current time. Depending on the power-on time of the field server, the time becomes a negative value; however, in this case, a value of 3,600 is added.

$$t_2 = t_1 - 30 \cdot \mathrm{FSID} \qquad (2)$$

The FSID can be used in the range of 0x00 to 0x77, and the field servers are started in the ascending order of FSID. By using FSID, simultaneous activation of each field server is prevented, and transmission signals of the field server are prevented from colliding. It became possible to transmit once every 30 seconds and 120 field servers are able to connect one master unit. Based on the results of experiments in the field, 30 seconds was deemed appropriate.

The PIC microcomputer controlling the power ON/OFF circuit always counts 0 to 3,599 s with the internal clock. When the time within the PIC microcomputer reaches 3,600 (0) s, power is supplied to the AVR microcomputer. Upon receiving the time correction signal from the master unit, the field server corrects the time within the PIC microcomputer to the correction time transmitted from the master unit and continues counting. This leads to the time synchronization between the field server and master unit. Even during the second and subsequent runs, when the time within the PIC microcomputer reaches 3,600 (0) s, power is supplied to the AVR microcomputer.

If the master unit fails to normally receive data from the field server due to a communication error etc., the master unit maintains the reception standby state of the sensor data without transmitting the time correction signal.

6 EXPERIMENTAL RESULTS

6.1 Consideration of Communication Protocol

For rice farming, affordable pricing is the most important requirement, and thus, the proposed method is effective. However, the device must also be able to deal with time variations lasting as long as several seconds. In consideration of the variations, we set the length of one frame to approximately two times the required length. By doubling the frame length, we can avoid collisions due to variations and secure time to retransmit. Increasing the frame length reduces the efficiency; however, the reliability of the system is improved.

LoRaWAN is a protocol that can be used for multiple purposes. Applications that collect meter reading values of gas, electricity, and water supply and applications like rice cultivation use the same protocol, although the communication frequency, communication time and intervals are different. In pursuit of price and performance, general-purpose functions are often wasted. For example, in the case of LoRaWAN Class A, the field server can send the data to the master unit at an arbitrary time, but in the case of rice field management, since data is acquired and transmitted at the same time, collision occurs frequently. In addition, although the master unit that receives data from the field server is a specification that transmits data to the field server twice, whether it is necessary to send this data twice depends on the application.

In the proposed protocol, we adopted scheduling to prevent conflicts. This is important for acquiring data at the same time and is an advantage of the proposed protocol.

6.2 Verification of Communication Protocol

We conducted the 7-day operation test to confirm whether the proposed communication protocol works as expected between the master unit and field server. Following are the points for the verification:

1) Confirm whether the master unit can return the time correction signal to the field server within the reception waiting time of the field server.
2) The field server that received the time correction signal confirms whether to shift to the sleep state immediately.
3) Confirm whether each field server properly changes the timing according to the time correction signal and starts at the specified time.

In the verification of the communication protocol, we used seven field servers, which is the same number used in our field. The distance between the field server and the master unit is centered on the master unit and all field servers are installed within a radius of 1 m. The sensor data sent from the field server to the master unit was saved in the verification cloud.

Table 5: Verification results of communication protocol.

Classification	Number of communications
Send sensor data	1,185
Number of resending	9
Completion of time synchronization (resend 1 time)	9

Table 5 shows the verification results of the communication protocol. The number of operational days is seven. The field server gets sensor data at an hourly interval, which is further sent to the master unit. The field server successfully sent the data 1,185 times. Of the 1,185 times, only nine were retransmission. Even when the first communication failed, reception of sensor data was successful from all field servers through retransmission. We confirmed that the protocol works for seven days without problems.

We implemented the designed communication protocols and carried out the operational test for a period of two months in an actual field. Figure 6 shows the measurement result at point C. A master unit was installed at point A. In this experiment, we confirmed that environmental data can be acquired every hour. The field server was equipped with sensors that can measure temperature, humidity, water level, soil temperature, and soil moisture content. The height at which the field server was installed was approximately 1 m from the ground surface to accurately measure the temperature. The height at which the master unit was installed was set to approximately 0.5m. It was confirmed that the measurement can be performed without problems, and data can be transmitted to the master unit. Owing to the fact that there is a communication failure at the rate of approximately 15.8%, the time correction may not be performed. The time synchronization was carried out when the fault was solved and it was confirmed that the protocol was operating properly. We examined the difference between the assumed startup time of the field server and the actual startup time. The results are shown in figure 17. The result displays the representative pattern of eight days from the operational test of two months. From this result, it is understood that when the time correction is performed, the error is suppressed to about in tens of seconds. Moreover, it is understood that the error is suppressed to 0 s in most communication between the master units to field servers.

The time error of the PIC microcomputer is the about 10 s in an hour from actual measurement [11]. This error is accumulated without time synchronization; however, in the field management communication protocol proposed here, this error is within the range in which collision with other field servers does not occur. Therefore, it is confirmed that time synchronization is effective in this communication protocol, and it is possible to reduce the increase in time error, which is proportional to the usage time. As a result, it became possible to transmit once every 30 seconds, and became possible to connect 120 field servers to one master unit.

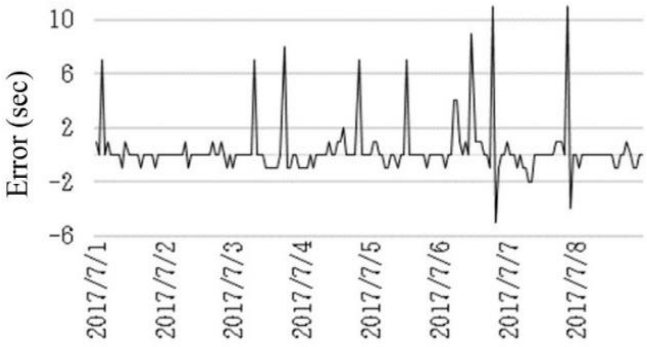


Figure 17: Time error of field server.

6.3 Evaluation of Power Consumption

Apart from the verification of the communication protocol, we conducted an experiment to verify the power consumption. The purpose of the verification is to obtain the power consumption during the operation. First, we measured the voltage, current value, and processing time for each operation mode.

The current measurement method is explained herein. In the case of a communication device, the current value varies depending on the communication state. Therefore, we decided to measure the current value while actually setting it in the field. In the rice field, there was no power supply, and thus, it was difficult to use a commercially available measuring instrument. Therefore, a current measuring device operating with a battery was developed, as shown in figure 18. INA 219 [12] was used for the sensor. In this measurement, the current was measured at intervals of 0.5 s and the change in current was observed. Based on the result, the change time to each mode and the average current in the mode were calculated. Figure 19 shows a field server with an ammeter is actually installed in the field.

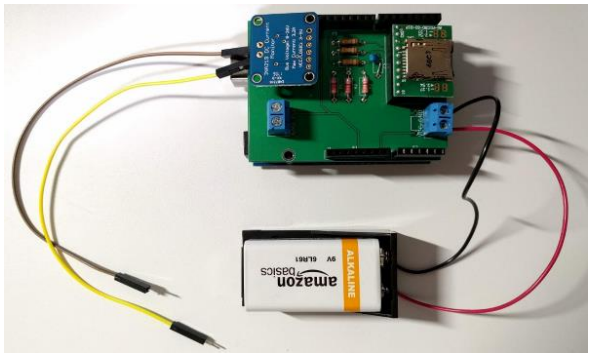


Figure 18: A field server and the sensors with battery.

Figure19: A field server and ammeter.

Table 5: Measurement results of power consumption.

	Time(s)	Current(mA)	Voltage(V)
Standby・acquire	6.55	45.8	5
Data send	1.65	86.6	5
Mode switching	3.9	50.1	5
Receiving standby	0.9	86.6	5
Receive	49	39.7	5
Sleep time	3538	0.167	5

Table 5 shows the measured results [13]. The data transmission mode is the most power consuming. It can be confirmed that the sleep time mode has the lowest power consumption among all.

Next, we calculated the power consumption and number of working days. Equation (3) shows the power consumption W [mWh] . Here, V_1 is the rated voltage [V] of the field server system. I_a is the electric current [mA] during the sensor stabilization standby and the sensor acquisition. t_1 is the time[s] during the sensor stabilization standby and the sensor acquisition. I_b is the electric current [mA] during the transmission of sensor data. t_2 is the electric time [s] during the transmission of sensor data. I_c is the electric current [mA] during the mode switching. t_3 is the time [s] during the mode switching. I_d is the electric current [mA] during the standby reception. t_4 is the electric current [mA] during the standby reception. I_e is the electric current [mA] during the data reception. t_5 is the time[s] the during the data reception. I_g is the electric current [mA] during the system sleep state.

$$W = (V_1\{(I_a \cdot t_1) + (I_b \cdot t_2) + (I_c \cdot t_3) + (I_d \cdot t_4) + (I_e \cdot t_5)\} + V_1 \cdot I_g\{3600 - (t_1 + t_2 + t_3 + t_4 + t_5)\}) / 3600 \quad (3)$$

We derived the number of operating days theoretically. In the case of retransmission is not occurred, the power consumption is 4.52 mWh per hour; the consumption being 108.4mWh per day. Therefore, theoretically, the field server can operate for approximately 691 days, assuming the electric quantity of the portable battery charger to be 75000 mWh. From the 7-day operation test described in table 5, 9 times of retransmission occurred by 7 field servers in 7 days, and retransmissions on the second times never occurred. If we assumed that one retransmission would necessarily occur with one transmission, the field server can operate for approximately 659 days.

Although the number of operating days has the theoretical value, it seems that the field server is able to operate for six months, which is the requirement of the agricultural corporation. Because rice field is softer soil, if you install a heavy field server it will fall over with wind etc. Therefore, it is important to reduce the weight of the battery, which is the heaviest component in the field server. From the experimental results in the rice field it has been found that it is necessary to reduce the number of D size battery to 3 or less. Therefore, 75000 mWh or less was set as the criterion.

Figure 20: A field server system in a rice field.

We conducted the operational test in actual rice fields using the 7 field servers of figure 6. The picture of the field server system installed in the rice field is shown in figure 20. We confirmed that the field server system works correctly from rice planting to rice reaping.

In IEEE 802.15.4e [14], two types of time synchronization methods, Beacon and Channel Hopping are defined. In the time synchronization defined in both methods, it is required that all nodes belonging to the network always synchronize the time within an error of ±1 ms, thereby realizing the time division access method. On the other hand, in the proposed method, time synchronization is performed between the master unit and each field server, but time synchronization between the field servers is not performed. Therefore, time synchronization accuracy of about ±10 seconds is sufficient, it is not necessary to hold hardware for special time synchronization and it is easy to put into practical use.

7 CONCLUSION

We proposed a new communication protocol, constructed a local wireless network, and conducted the experiment. In the field servers for the rice field using the LPWA technology, which require only batteries for operation, the proposed time synchronization is an important technology for the purpose of reducing the power consumption. Additionally, the proposed time synchronization is an important technique for increasing the line use efficiency. It was seen from the experimental results that the power consumption of the field server is 108.4mWh per day. Therefore, it was confirmed that the method can continuously work for 691 days based on our calculations. We confirmed that the field server system works correctly from rice planting to rice reaping. The time synchronization is effective and was able to decrease the timing error in direct proportion to the operating time. This protocol is valid for the rice cultivation management systems because the field server is stable and can operate for a long time. Therefore, it meets farmers' expectation to utilize a reasonable field server.

REFERENCES

[1] Ministry of Agriculture, Forestry and Fisheries of Japan, Statistics of agricultural labor, Accessed on 2017-6-2. [Online]. Available: http://www.maff.go.jp/j/tokei/sihyo/data/08.html.

[2] Kiyokazu Kurosawa, Isamu Iizima, Yoshiki Amemiya, Shunya Yamamichi, Masaharu Toyota and Mikiko Sode Tanaka, "Development of Operational Control System for Rice Cultivation Equipped with Activity History Function," IEICE Technical Report, vol. 116, no. 346, CS2016-55, pp. 59-64 (2016).

[3] Yuta Kawakami, Masaharu Toyota, Keitaro Terada, Keiko Matsumoto and Mikiko Sode Tanaka, "A study of the optimal agricultural field communication using Sub-GHz wireless technology," IEICE Technical Report, vol. 116, no. 382, NS2016-138, pp. 107-112.

[4] Hao Guo and Peter Crossley "Design of a Time Synchronization System Based on GPS and IEEE 1588 for Transmission Substations," IEEE Transactions on power delivery, vol. 32, no. 4, (2017).

[5] S. Ganeriwal, R. Kumar and M. B. Srivastava, "Timing-sync Protocol for Sensor Networks," Proceedings of the 1st ACM Conference on Embedded Network Sensor Systems (SenSys'03), Los Angeles, California (2003).

[6] Paddy watch, Accessed on 2018-10-2. [Online]. Available: https://field-server.jp/paddywatch/rental/index.html

[7] Sigfox, Accessed on 2018-10-2. [Online]. Available: https://www.sigfox.com/en

[8] IEEE 802.15.4 Working Group, Accessed on 2018-10-2. [Online]. Available: http://standards.ieee.org/develop/project/802.15.4.html

[9] ZigBee alliance, Accessed on 2018-10-2. [Online]. Available:http://www.zigbee.org/

[10] Masaharu Toyota, Keitaro Terada, Yuya Takada, Tadaaki Hirata, and Mikiko Sode Tanaka, "Construction of rice cultivation management network with LoRa," IEICE Technical Report, vol. 117, no. 3, NS2017-11, pp. 61-66, 2017.

[11] Keitaro Terada, Masaharu Toyota, Tadaaki Hirata, Yuya Takada, Keiko Matsumoto and Mikiko Sode Tanaka, "Proposal of communication protocol for field management using LoRa," DICOMO2017, pp. 1671-1678, 2017/6/28-30.

[12] INA219 High Side DC Current Sensor, Accessed on 2018-10-2. [Online]. Available: https://www.adafruit.com/product/904

[13] Yumeto Kojima and Mikiko Sode Tanaka, "Current value measuring device for field server of field management using LoRa," IEICE Society Conference 2018, 2018/9/11-14.

[14] IEEE Standard for Local and metropolitan area networks--Part 15.4: Low-Rate Wireless Personal Area Networks (LR-WPANs) Amendment 1: MAC sublayer Accessed on 2018-10-2. [Online]. Available: https://ieeexplore.ieee.org/document/6185525/.

(Received October 19, 2018)
(Revised December 4, 2018)

Koichi Tanaka received B.E., and M.E. degrees in Information Science and Technology from Kanazawa Institute of Technology in 1985, 1987. His research interests include mobile computing, distributed systems and telecommunication protocols such as field servers for cultivations, car navigation systems, and mobile phones. He is a doctoral student of Shizuoka University from 2009. He is a member of IPSJ (Information Processing Society of Japan).

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

Tomochika Ozaki received the B.E. degree from the Nagoya University in 1988, the M.E. degree from the Nagoya University in 1990 and received the Ph.D. degree in Informatics from Shizuoka University, Japan, in 2018. In 1990, he joined Hitachi Ltd. His research interests include embedded systems, energy management systems and human machine interface. He is a member of IPSJ (Information Processing Society of Japan).

Masakatsu Nishigaki received his Ph.D. in Engineering from Shizuoka University, Japan. He served as a Postdoctoral Research Fellow of the Japan Society for the Promotion of Science in 1995. Since 1996 he has been engaged in research at the Faculty of Informatics, Shizuoka University. He is now a Professor at the Graduate School of Science and Technology of Shizuoka University. His research interests are in wide variety of information security, especially in humanics security, media security, and network security. He served as Chief Examiner of IPSJ (Information Processing Society of Japan) Special Interest Group on Computer Security from 2013 to 2014, Chair of IEICE (Institute of Electronics, Information and Communication Engineers) Technical Committee on Biometrics from 2015 to 2016, and currently serving as Director of JSSM (Japan Society of Security Management) since 2016. He is IPSJ (Information Processing Society of Japan) fellow.

Tadanori Mizuno received the B.E. degree in Industrial Engineering from the Nagoya Institute of Technology in 1968 and received the Ph.D. degree in Engineering from Kyushu University, Japan, in 1987. In 1968, he joined Mitsubishi Electric Corp. From 1993 to 2011, he had been a Professor at Shizuoka University, Japan. From 2011 to 2016, he had been a Professor at the Aichi Institute of Technology, Japan. Since 2016, he is an Affiliate Professor at the Aichi Institute of Technology, Japan. His research interests include mobile computing, distributed computing, computer networks, broadcast communication and computing, and protocol engineering. He is a member of IPSJ (Information Processing Society of Japan), IEICE (Institute of Electronics, Information and Communication Engineers), the IEEE Computer Society and Consumer Electronics, and Informatics Society.

Regular paper

Unsupervised Biometric Anti-spoofing using Generative Adversarial Networks

Vishu Gupta†, Masakatsu Nishigaki†, and Tetsushi Ohki†

†Faculty of Informatics, Shizuoka University, Japan
vishu@sec.inf.shizuoka.ac.jp, {nisigaki, ohki}@inf.shizuoka.ac.jp

Abstract - With the advent of new technologies, the methods of presentation attacks as well as the security measures taken against it is diversifying with each passing day and are competing with each other. The imposter can make access to a system illegally by deceiving the security through the use of material containing artificial biometrics traits like a printed photo, display, etc. Therefore, we propose a novel presentation attack detection algorithm which can ensure security against unknown presentation attacks without any prior knowledge of fake samples. Moreover, our proposed algorithm can detect presentation attack with a single static image only. The essential tasks are divided into two parts, creating a smooth manifold of live samples and determining whether the manifolds includes the query image. In this paper, we utilize one class system such as SVM(Support Vector Machine) and DCGAN(Deep Convolutional Generative Adversarial Network) to learn the manifold of live samples. For DCGAN we propose a liveness scoring scheme based on the AnoGAN(Anomaly Generative Adversarial Network) Framework. Based on these, we utilize the proposed method to palm presentation attack detection. Through our experiment, we were able to produce decent results by using palm live/fake image dataset.

Keywords: biometrics, spoofing, presentation attack detection, anomaly detection, generative adversarial networks

1 INTRODUCTION

Along with the development of artificial intelligence and cryptographic technology, a society approaching not only simple tasks but also decision making of people to computers is coming. In such a society, it will become an important requirement to guarantee that the outsourcing was performed by the user's own will, and also to correctly detect it when counterfeiting acts are forged or improperly tampered. It is essential to guarantee the authenticity of the terminal in addition to the authenticity of the terminal itself to satisfy these requirementsṪhe biometric authentication system is drawing attention, which can guarantee the authenticity of the terminal user.

Biometric authentication system (BAS) registers preliminary collected biometric information as a template and verifies whether it belongs to a legitimate user by calculating the similarity with the biometric information acquired at the time of authentication. BAS uses a biometric feature of the person without fear of forgetting, losing, or theft compared to an authentication method using a password or a token. In addition to the advancement of traditional application in fields such as immigration control, ATM, the entry and exit management, recent years, personal use in mobile terminals has been expanding.

On the other hand, biometric information such as faces, sounds, fingerprints, handwriting is difficult to keep secret in daily life. Biometric presentation attack is becoming a significant threat since false biometric information becomes more sophisticated along with the rapid development of sensors, printers, and manufacturing machines.

To develop a BAS that is secure against presentation attacks, demand for designing a robust presentation attack detection(PAD) algorithms which classify an input sample as live or fake is increasing.

Many previous approaches discussed the PAD features which can guarantee security against a specific impersonation attack such as frequency spectrum for printed photo [1], [2], three dimensionalities of live face [3], motion-based feature for video [4] and so on. However, the methods of presentation attacks are diversifying day by day. It is difficult to learn in advance PAD features that can detect all these attacks.

Regarding the problem, PAD algorithms have made it possible to detect various presentation attacks by combining multi-class classifier that solves the classification problem between live and various fake samples such as [5]–[8]. However, these methods still have some issues. At first, it is necessary to obtain not only biometric samples but also a large number of fake samples for each type of presentation attack. Second, the PAD algorithm does not guarantee whether an anomaly sample is classified as a presentation attack. Here we define anomaly sample as a sample that is not included in the samples for training. Note that anomaly sample includes not only samples intended to resemble a live sample but also any synthetic samples since it is sufficiently effective in the registration process. There exists an attack using synthesized input that can impersonate the majority of registered users [9]. Also, attacks that send arbitrary commands to unregistered home interactive speakers by using sounds in the inaudible area [10]. Capturing such attacks with pre-trained PAD features is difficult.

The subject of this paper is to investigate the security against the presentation attack using an anomaly sample as features of biometric information such as the face, palm, etc. differ depending on the modality. Therefore, we utilize DCGAN to perform the estimation of the distribution of biometric information to solve the fundamental problem of making PAD difficult due to the diversification of attacks. Moreover, it is impossible to predict the counterfeit that will be used as an

impersonation of the real sample while performing the anti-spoofing using a biometric system. Therefore, by making use of one class system neural network which uses anomaly detection to distinguish fake sample from the true sample, we can make a better system to counter spoofing attacks.

In the experiment conducted in this paper, we used a custom-made database created out of palm images as it was easy to create unknown samples as well as it was found realistic that the attacker may perform the counterfeit attack by using a rubber glove, displayed photo, etc. in an attempt to break into the system.

Additionally, our method relies on a single static image to detect presentation attack. Such a method can also be directly applied to deal with video spoof or be integrated with a video-based palm PAD algorithm for better performance. The main contributions of this work are as follows:

1. We propose a novel Presentation Attack Detection (PAD) algorithm which can be learned only with live samples and guarantee security against an anomaly sample by utilizing GAN based anomaly detection algorithm.

2. Proposed PAD algorithm is evaluated with custom-made database (Custom-Made Database containing live and fake palm samples) and achieved 3% of HTER (Half Total Error Rate) by using a model trained only with live samples.

2 RELATED WORK

All the prior research that has been conducted on Anomaly detection is performed by having to train the system by using both live samples along with fake samples which are used for presentation attack. For all these conducted researches, the core difference lies in the method used to model the real and fake attempts. Prior methodologies based on the employed cues are being classified in a recent study [11] where they are divided into three major categories.

The first category is a method to detect face liveness which relies on image quality/distortion measures. Work in [12] which consists of identifying print attacks using the difference in the 2D Fourier spectra is an example of the method in this category. The work stated in [13] utilizes the Lambertian model which comprises of variational Retinex based approach and Gaussian filters difference as its two methods. The work in [14] uses power spectrum and local binary patterns [15] to exploit both frequency and texture information. [16] modeled spatial and temporal information for face presentation attack. [17] proposes the combination of motion and texture methods via score level fusion. Difference-of-Gaussian filters to choose specific frequency bands for feature extraction was done in [18]. The work done in [19] proposes presentation attack detection by analyzing the texture represented using multi-scale local binary pattern [15] which provides a unique feature space for coupling spoofing detection and face recognition. The results from [20] reported good performance on Replay-Attack database.

The second category uses methods which are based on detecting different signs of vitality which make use of characteristics corresponding to live faces. For example, presentation attack detection in [21] uses blinking which is used with others cues in other work. Such as [22] recommend the use of all the dynamic information content of the video represented using dynamic mode decomposition method. The work done in [23] utilizes both eye-blink and scene content clues as a hybrid face liveliness detection system against spoofing with photographs, videos, and 3D models of a valid user in a face recognition system.

The last category consists of methods based on the difference in motion patterns between real and presentation attacks. It is assumed that the presentation attack have rigid motion whereas real-access attempts has both rigid and non-rigid motion. This approach depends on the fact that real accesses correlate with 3D structures whereas presentation attack media are often at 2D planes. Eulerian motion magnification using two sets of features composed of LBPs(Local Binary Patterns) [15] to enhance facial expressions is a typical case of the method in this category. The new liveness detection method is proposed in [24] which utilizes the difference in optical flow fields generated by the movements of 2D planes and 3D objects. A countermeasure against face presentation attack was proposed in [4] which were based on foreground/background motion correlation using optical flow showing promising results on the Photo-Attack database. The work in [3] used geometric invariants to detect replay attacks once a set of automatically located facial points are detected which was evaluated on two publicly available databases of NUAA [13] and HONDA [25].

While most of the existing methods use real access data to try and learn a general classifier to outline presentation attack attempts, work in [26] uses both texture and motion cues, the authors built two presentation attack detection methods, one being a generative approach while the other being a discriminative method to study the client-specific information embedded in the feature space and its effects on the performance of the system. Similarly, the work in [27] proposes a method using a classifier trained explicitly for each subject.

The current work regarding detection mechanism share some similarities to the existing approach which utilizes image content representation is distinct in the way we formulate the existing the detection problem. The standard approach used to detect an anomaly in an image uses two-class formulation where they separate the negative from the positive samples, our proposition uses one-class pattern classification methods, testing it in a modified as well as an existing method which yields good results to identify presentation attack attempts. Moreover, the evaluations are performed by using a custom-made database which better reflects the difficulties of detection in realistic scenarios. Also, many of the existing papers are supervised and conducted using face images/videos. These papers are evaluated using public databases such as Replay Attack Database. However, all of these public databases aims at the evaluation of counterfeit samples that imitate living organisms and does not assume the possibility of attacks that are performed by using unknown samples.For this reason, in this paper, we created our custom-made database of palm for evaluation by considering the possibility of various unknown samples. The reason for choosing Palm as a modal-

ity is that it is smaller in size as compared to face, the database can be made easily with an inexpensive camera, and it is easy to create an unknown counterfeit of the entire palm by wearing gloves or by making a false palm out of different compounds.

3 PRESENTATION ATTACK DETECTION USING ANOMALY GAN

3.1 Generative Adversarial Networks

Goodfellow et al. introduced a concept of Generative Adversarial Network(GAN) [28] which learns a *generator* expression indistinguishable by a *discriminator* by training a *generator* model and *discriminator* model simultaneously. The aim of the *generator* is to fool the *discriminator* by learning the probability distribution of the input samples. Let $\boldsymbol{x}$ be an input sample whose true probability distribution is $p(\boldsymbol{x})$. G is a *generator* that takes a latent vector $\boldsymbol{z}$ randomly selected from the latent space $\mathcal{Z}$ and outputs a new sample $G(\boldsymbol{z})$. The *discriminator* D then outputs the probability that the given input is either the true input from $p(\boldsymbol{x})$ or the $G(\boldsymbol{z})$ from the *generator*. These two models are simultaneously trained using the min-max game of the formula:

$$\min_D \max_G V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))] \quad (1)$$

Radford et al. [29] introduced deep convolutional generative adversarial networks (DCGAN) for unsupervised learning of features by utilizing convolutional neural networks as the *generator* and *discriminator* network. More specifically, they replaced the pooling layer with stride convolution layer so that the network can learn its own spatial upsampling. Additionally, they removed the full connection layer at the top of the convolution feature to improve model stability. Finally, batch normalization was utilized to suppress training problems caused by poor initialization and helps the propagation of gradients in deep models by normalizing each unit to have zero mean and unit variance.

3.2 Proposed Anomaly GAN for PAD

To detect presentation attack using a single image, we propose unsupervised learning to identify anomalies in imaging data as candidates for the fake sample. Fig. 1 shows an overview of our proposal. Our proposed scheme is based on unsupervised anomaly detection scheme proposed in [30] which is aimed at detection of disease markers in medical imaging(hereafter, AnoGAN). AnoGAN uses DCGAN to learn a manifold of live sample variability, accompanying an anomaly scoring scheme based on the mapping from image space to a latent space.

3.2.1 Palm Imaging Model

We learn the palm image manifold $\mathcal{X}$ on the image space with unsupervised learning using only the live palm images. When a query image is not included in the learned manifold $\mathcal{X}$, it can be detected as an unknown input.In DCGAN [29], *generator* uses latent vector $\boldsymbol{z}$ chosen from latent space $\mathcal{Z}$ uniformly at random to obtain a smooth mapping $G(\boldsymbol{z})$ to palm image manifold $\mathcal{X}$.

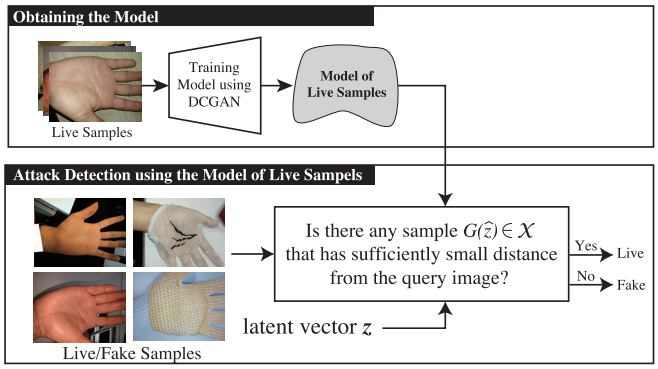


Figure 1: Overview of our proposal.

3.2.2 Deriving Latent Vector

We can detect the Presentation Attack by checking whether query image $\boldsymbol{x}_q$ is included in the palm image manifold $\mathcal{X}$ learned in the clause 3.2.1. Since DCGAN calculates $G(\boldsymbol{z})$ using the randomly chosen latent vector $\boldsymbol{z}$, $G(\boldsymbol{z})$ corresponds to a random point on the palm image manifold $\mathcal{X}$. Consequently, the distance between $G(\boldsymbol{z})$ and the query image $\boldsymbol{x}_q$ does not necessarily become small even if it is a live sample. Therefore, to detect an anomaly sample, we should confirm the existence of a latent vector $\widehat{\boldsymbol{z}}$ that has a sufficiently small distance between the query image $\boldsymbol{x}_q$ and $G(\widehat{\boldsymbol{z}})$ on the manifold $\mathcal{X}$.

For finding the $\widehat{\boldsymbol{z}}$ from randomly chosen latent vector $\boldsymbol{z}$, we use the backpropagation approach proposed in [30]. The loss function $\mathcal{L}(\boldsymbol{z}_\gamma)$ for backpropagation is defined as follows:

$$\mathcal{L}(\boldsymbol{z}_\gamma) = (1-\lambda)\cdot\mathcal{L}_R(\boldsymbol{z}_\gamma) + \lambda\cdot\mathcal{L}_D(\boldsymbol{z}_\gamma) \quad (2)$$

where $\boldsymbol{z}_\gamma$ is an updated latent vector to fool *discriminator* D, $\mathcal{L}_R(\boldsymbol{z}_\gamma)$ is the generator loss, $\mathcal{L}_D(\boldsymbol{z}_\gamma)$ is the discriminator loss and λ is a fixed parameter for convex combination. The residual loss and the discriminator loss can be obtained as follows:

$$\mathcal{L}_R(\boldsymbol{z}_\gamma) = \sum |\boldsymbol{x}_q - G(\boldsymbol{z}_\gamma)| \quad (3)$$

$$\mathcal{L}_D(\boldsymbol{z}_\gamma) = \sum |\boldsymbol{f}(\boldsymbol{x}_q) - \boldsymbol{f}(G(\boldsymbol{z}_\gamma))| \quad (4)$$

where $\boldsymbol{f}(\cdot)$ is an output of the *discriminator* function. Only the coefficients of $\boldsymbol{z}$ are adapted via backpropagation. The trained parameters of the *generator* model and *discriminator* model are kept fixed. In our proposal, $\widehat{\boldsymbol{z}}$ is obtained by applying backpropagation process α times with query image $\boldsymbol{x}_q$ and randomly selected $\boldsymbol{z}$. The obtained $\widehat{\boldsymbol{z}}$ is used in classification process.

3.2.3 Classification

In classification process, we investigated the three types of score function, anomaly score $A(\boldsymbol{x})$, residual score $R(\boldsymbol{x})$, and

discriminator score $D(\boldsymbol{x})$, respectively. The relationship between each score is defined as follows:

$$A(\boldsymbol{x}_q) = (1-\lambda)\cdot R(\boldsymbol{x}_q) + \lambda \cdot D(\boldsymbol{x}_q) \tag{5}$$

where the residual score $R(\boldsymbol{x}_q)$ and discrimination score $D(\boldsymbol{x}_q)$ are defined by the residual loss $\mathcal{L}_R(\widehat{\boldsymbol{z}})$ and discriminator loss $\mathcal{L}_D(\widehat{\boldsymbol{z}})$ using at the α update iteration of the mapping procedure to the latent space, respectively. All score functions output a large score for an anomaly image. In our experiments, we use λ = 0.9 in equations (2) and (5) which was found empirically due to preceding experiments on our palm dataset.

3.2.4 Cumulative score calculation

The experiment performed in [30] requires the trained model to be executed for α times and also requires to perform numerous backpropagation steps even if the sample was obviously an anomaly sample. Therefore, We propose to utilize a cumulative score $C(x_q, \beta)$ for a query image x_q at β-th backpropagation step which is as follows:

$$C(x_q, \beta) = \sum_{b=0}^{\beta} A(x_{q,b}) \tag{6}$$

where $A(x_{q,b})$ is an anomaly score for b-th backpropagation step.

If the target is a live sample, $A(x_{q,b})$ will decrease more sharply as the backpropagation step increases since $G(z)$ and live sample are within the same manifold $\mathcal{X}$. Therefore, if we assume that α is the maximum count for the execution of backpropagation, then the input sample can be classified as a live sample if the value of cumulative score $C(x_q, \alpha)$ is smaller than the threshold th. On the other hand, if at a certain point β whose value is $\beta < \alpha$, the calculation can be canceled and the input sample can be classified as a fake image as soon as the cumulative score satisfies $C(x_q, \beta) > th$. For this reason, it is possible to reduce the amount of calculation as compared with the usual method which always requires α times calculation for the backpropagation.

4 EXPERIMENT

In this section, first, a description of the custom-made database and the evaluation protocols used in this experiment is provided, following by experimental results obtained from the database used. All the experiments were carried out using Python with the tensorflow and pytorch library on a machine with configuration (Intel i7-5930K, 64GB RAM, 12x Intel(R) Core(TM), Ubuntu 64bit) environment.

4.1 Database

Many previous works have used public live/fake dataset such as Replay-Attack Database [2] and Unconstrained Smartphone Spoof Attack (USSA) Database [31]. However, it contains only a specific type of fake photo and video samples making it inadequate in terms of anomaly samples. Therefore, in our experiment, we constructed a custom-made database to make sure that the system is being able to make a clear distinction between live and fake samples even when the system encounters unexpected inputs such as palm with a glove, palm with a vinyl glove, etc. which have no direct relation with the hand. So, in our custom-made database, we prepared a large amount of data to check whether the system will be able to counter any fake sample provided to it by the attacker as an input.

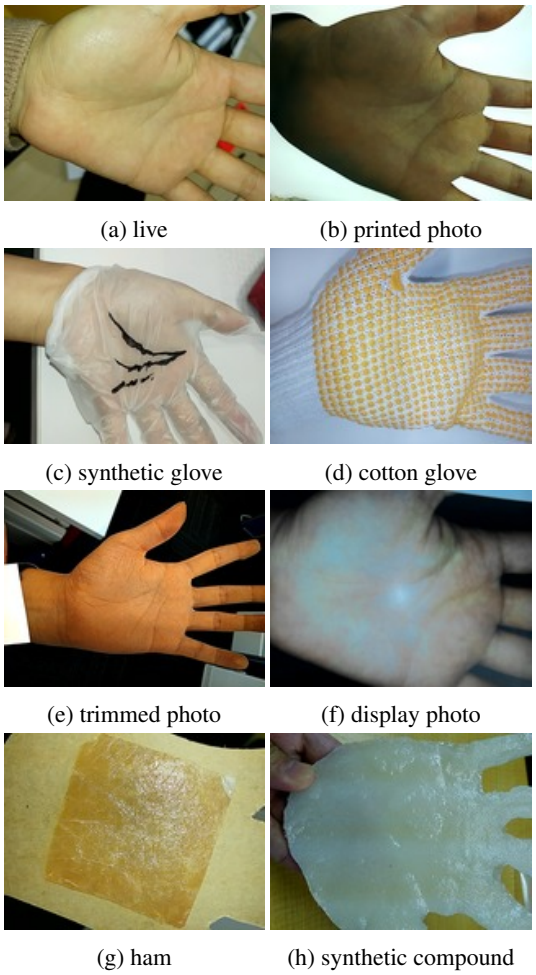
The custom-made database used in the experiment consists of 8748 live samples and 6648 fake samples of palm with an image resolution of 160x120 pixels taken directly from approximately 2000 people with ten different types of mobile cameras (LG G5, LG Nexus 5x, LG Nexus 5, Sony Xperia X Performance, Elephone P9000, Sharp Aquos SHV34, Doogee X5max, Huawei GR5 (KII-L22), ASUS zenfone2 (Z00D), ASUS P008). The images are taken in different non-controlled indoor surrounding conditions such as, inside office with different background or inside the building with varying conditions of lighting which also includes photo that is made in a dark place with the help of flashlight ,etc with varying postures in order to anticipate all kind of possibilities of the images that will be used as the input for the system. The training set used to train the AnoGAN model comprises of randomly selected 8000 live samples. The test set in total consist of 7396 samples out of which 748 were live palm samples and 6648 were fake palm samples from cases not included in the training set. The training that we are performing in this experiment is uncontrolled without of external interference. Example of true samples and different variety of fake samples that were used while training the system is given as below in Fig. 2. In order to include as many variety of unexpected fake samples as possible to check the accuracy of the system, we included photos such as (b)printed photo, (c)hand wearing synthetic glove, (d)hand wearing cotton glove,(e)printed photo that were cut from the border of the hand part in order to resemble hand in 2D, (h)gelatin or (g)ham which may not have direct relation with the hand but can resemble skin and (f)photo that were taken from a digital device such as iPad or webcam.

4.2 Evaluation Protocol

The manifold of live images was solely learned on image data of live cases with the aim to model the variety of live appearance. So for that purpose, 8000 live samples are selected from the database as noted earlier to develop the model. In a real case scenario, it is difficult for users to collect 8000 images in order to create a model for such evaluation, but this problem can be solved by using learned models provided by vendors who have easy access to a large amount of data which will be used to generate the required model for the experiment, real-life use, etc. For performance evaluation in anomaly detection, we ran various protocols exploited by researchers.

4.2.1 Our Protocol

All the training and test conducted for the anomaly detection in this work are based on the one class system where only the



(a) live (b) printed photo

(c) synthetic glove (d) cotton glove

(e) trimmed photo (f) display photo

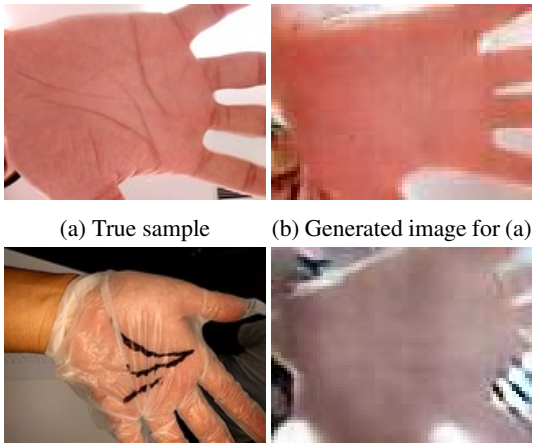
(g) ham (h) synthetic compound

Figure 2: Example samples used for training the model for 1 class system. (a) is an example of true sample used for training the model and (b) to (h) are the different variety of fake samples that were used for testing the model.

live samples are used to develop the model. In particular, the following systems are used for the development and evaluation:

- AnoGAN+RAW: The AnoGAN which uses one class system trained using the original image
- SVM1+RAW: The one-class SVM with a Gaussian kernel trained using the original image
- SVM1+LBP: The one-class SVM with a Gaussian kernel trained using the LBP feature

For each of these protocols, the model was trained using 8000 live samples and the test was conducted by using 100 fake samples along with 100 live samples which were not included in 8000 live samples that were used for training the model to check the accuracy of the model in order to distinguish the fake samples from the live samples. Residual score $R(\boldsymbol{x}_q)$ is taken into account in order to differentiate between live samples and fake samples. The purpose of this unsupervised one class training is to find the epochs whose training accuracy as well as prediction accuracy are good and which does not cause overfitting. For this dataset the epochs which



(a) True sample (b) Generated image for (a)

(c) Fake sample (d) Generated image for (c)

Figure 3: Example samples used for training the model for 1 class system and the respective image produced by the AnoGAN after performing 100 backpropagation. (a)True sample (b)Image generated by AnoGAN for true sample (c)Fake sample (d)Image generated by AnoGAN for fake sample.

showed the best result is 50. As we try to increase the epochs the accuracy of the model decreased. In this experiment, the unsupervised learning was conducted by changing the epochs as 20, 25, 50,$\cdots$,100. In the result section of this experiment, we used the result of 25 epochs as the representative example and the result of 50 epochs as it shows the best result. The residual score can vary each time the test is conducted even if the image used for testing is the same because the residual score measures the visual dissimilarity between query image $\boldsymbol{x}_q$ and generated image $G(\widehat{\boldsymbol{z}})$ in the image space by finding a point $\widehat{\boldsymbol{z}}$ in the latent space that corresponds to an image $G(\widehat{\boldsymbol{z}})$ that is visually most similar to query image $\boldsymbol{x}_q$ and that is located on the manifold $\mathcal{X}$. We ran 100 backpropagation steps (α = 100) for the mapping of new images to the latent space $\mathcal{Z}$. The image produced after performing 100 backpropagation is given as in Fig. 3 which shows that the trained model can generate reasonably realistic looking images when a live sample is used for the classification of the image as the image is generated from inside the manifold $\mathcal{X}$. On the other hand, when a fake sample is used for the classification process, as the images are obtained from the same manifold, images that are close to real to some extent can be obtained. However, as the query sample is unreal, the residual score gets bigger.

4.2.2 Evaluation Metric

For evaluating the result obtained, we consider the Area Under Curve (AUC) obtained from Receiver Operating Characteristic (ROC) curves. The ROC curve was made using the residual score as the parameter which yields good results as shown in [30]. The vertical axis and the horizontal axis of ROC curves usually present True Positive and False Positive Rate respectively. It indicates that the plot 's top left corner is the optimal point. Preferable TPR for the ROC curve is equal to one which makes the excellent AUC's values approaching one.

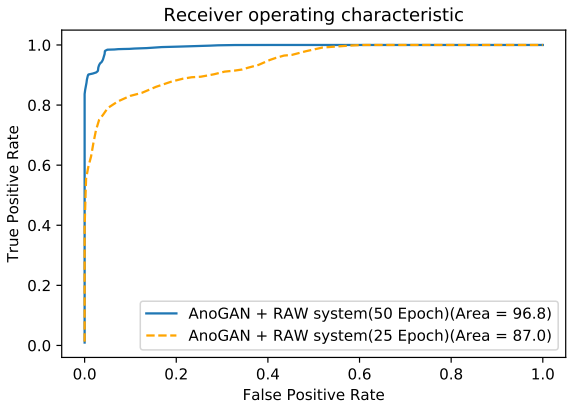


Figure 4: The above figure represents the ROC graph of the AnoGAN model trained for 25 (orange) and 50 (blue) epochs respectively by using live samples as an input image for training.

Table 1: Area under the ROC (AUC) (%) for different systems obtained by using custom database.

System	AUC(%)
AnoGAN+RAW (20 Epoch)	83.1
AnoGAN+RAW (25 Epoch)	87.0
AnoGAN+RAW (50 Epoch)	96.8
SVM1+RAW	34.3
SVM1+LBP	83.5
SVM1+LBP (cos similarity)	75.9

4.3 Evaluation Results

The one-class systems introduced earlier are evaluated on the custom-made database which used 8000 live samples to develop the model. To make sure that there is no bias in the result obtained after testing each of the models we took out a total of 200 samples randomly from the palm database, 100 samples each from live samples and fake samples. For the fake samples, even though the 100 images taken out were selected at random, it was made sure that it contained all the variety of samples that were taken into account while creating the fake samples. By doing so, we can see to what extent the trained model produces the desired result even if it encounters unexpected input which is fake but has no direct relation with hand.

Figure. 4 represents the ROC graph of the results obtained from different models where the Y-axis shows the True Positive Rate and the X-axis shows False Positive Rate. Additionally, Table 1 and 2 show the AUC and HTER (Half Total Error Rate) respectively. Note that HTER can be calculated by $\min(TN + FP)/2$.

4.3.1 Accuracy

Table 1 shows that the best performing one-class system regarding average performance is Ano-Gan+RAW with an average AUC of 96.8%. The result obtained regarding AUC by using one class SVM system [32] as a model for training and testing the dataset as used for training and testing AnoGAN is also shown in Table 1 and 2. It is clearly visible that the proposed AnoGAN system is far more better than the conventional one class SVM system. As far as the security check for AnoGAN is concerned, it can be conducted by using the image produced by the AnoGAN as the input image while testing the model and comparing the residual score with that of the unseen real samples and fake samples. As far as the one class SVM system are concerned, SVM1+LBP performed better as compared to SVM1+RAW. SVM1+LBP is more sensitive whereas SVM1+RAW is less sensitive to the attack model because as stated in [19] by using LBP feature they were able to perform their experiment in a robust way which was computationally fast and didn't required any user-cooperation. Moreover, the extensive experimental analysis done by them on a publicly available database showed excellent results compared to existing works which proves clarifies that SVM1+LBP will show better results as compared to SVM1+RAW.

Table 2: Half Total Error Rate(HTER)(%) for different sys-tems obtained by using custom database.

System	HTER(%)
AnoGAN+RAW (20 Epoch)	17
AnoGAN+RAW (25 Epoch)	13
AnoGAN+RAW (50 Epoch)	3
SVM1+RAW	34
SVM1+LBP	17
SVM1+LBP (cos similarity)	20

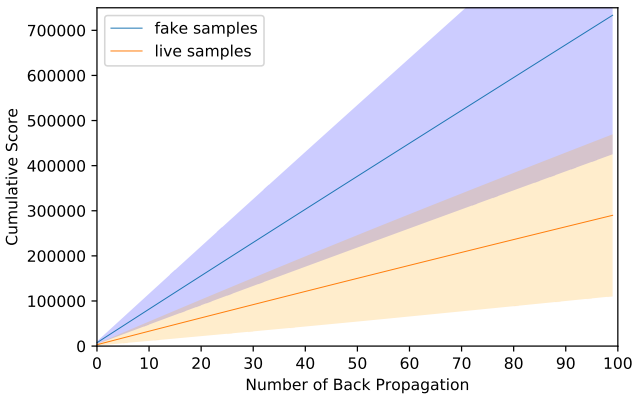


Figure 5: The above figure represents cumulative score calculation of the AnoGAN model trained for 50 epoch where the average of cumulative score for each epoch along with the standard deviation is taken into consideration.

4.3.2 Computational efficiency

As described in section 3.2.4, when using the cumulative score to perform the analysis, fake input can be detected at β backpropagation which is less than the maximum number of backpropagation α. At this time, in order to calculate the efficiency we use the ratio of the average number of backpropagation $\overline{\beta}$ and the maximum number of backpropagation α that

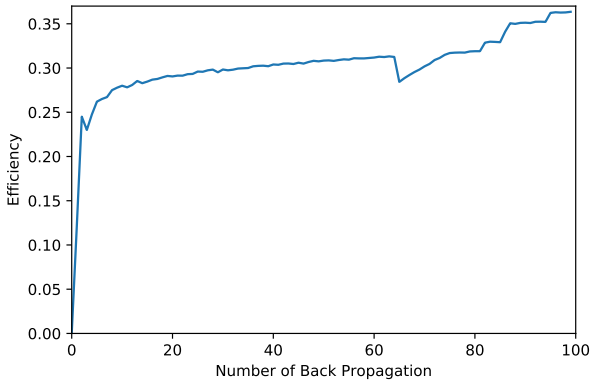


Figure 6: The above figure represents the possible efficiency of the AnoGAN model on the basis cumulative score at each backpropagation.

is $\overline{\beta}/\alpha$ as efficiency and evaluate the computational effectiveness of cumulative score.

Figure. 6 shows the value of efficiency when α is changed. The protocol mentioned in section 4.2 was used for training and evaluation of AnoGAN. Regarding the threshold th for each α, we calculated and applied the threshold value that minimizes HTER in test data. As it can be seen from Fig. 5, the larger the α, the greater the efficiency and the greater the effect of the cumulative score. Also, since efficiency > 0 is always valid except for $\alpha = 0$, it is confirmed that using cumulative score always has a computational advantage compared to using residual score.

5 DISCISSION

In the one-class system, the AnoGAN method is more accurate as compared to conventional one-class SVM which produced good results in other researches. Among the models that were produced by the AnoGAN system, the model which was trained for 50 epoch showed the better result as compared to other models. From the results we obtained by performing this, we can see that an unsupervised model such as AnoGAN could have many benefits, we also see some research limitations.

First, the number of epochs for which you have to train the system may depend on the number of images that you are using to train the system. However, we have not yet found out the relation between them. Therefore, finding the relationship and optimizing the number of epoch needed from training the model could help us improve our accuracy.

Second, Infinite samples can be produced from the DCGAN used in AnoGAN as it digitally produces images which are considered as a real sample by the system generating the model which can be used to improve the accuracy of the system and give better results while calculating the residual score for a given input. Here a doubt arises that, as the DCGAN that we are using while performing this experiment can generate an infinite number of samples, it can be thought that the attacker can misuse the produced image by using it to attack the system, but it is unlikely to happen. This is because, even when the result of the GAN is obtained as an image, it is necessary to output it into some other form to make it visible to us such as paper or a display and shoot with a camera. In this method, images of living organisms are produced onto a paper or a display, and samples that are photographed by cameras are detected as a fake sample with high probability. So, it is not possible to misuse the output of the DCGAN. On the other hand, as features of biometric information such as the face, palm, etc. differ depending on the modality, it was considered possible to perform the estimation of bio-distribution from different modality using DCGAN. However, it was not fully verified as we didn't had many unknown sample in our custom-made database. It is necessary to do the future experiment by including more unknown sample in the data set.

Third, While testing the one-class SVM method the SVM1 + LBP produced better results as compared to SVM1 + RAW. Therefore if the code of AnoGAN is designed in such a way that it calculates the loss function while taking LBP (histogram, cosine similarity) into consideration from the point of training, then there is a chance that it might produce a better result. Also, we have only examined anomaly detection systems based on 1 class SVM and AnoGAN; it would be better if we study other anomaly detection approach also.

Fourth, it can be concluded that even if you train the system by using the true samples only, it does not perform well enough and more modification and research should be conducted to improve the performance of this type of system. Therefore, to improve the performance, we would like to take the cost included in the making of the data set which would serve as a checkpoint from where we can strive for further improvement.

Last, in the experiment conducted this time, we have only used the image captured by using the camera that can be found in any of the typical mobile used by us in our day to day life. We also did not control or limit the posture of the palm at the time of the shooting. This is done to increase the robustness of the model as we assume that by doing so the system will be able to adjust its identification analysis concerning the user's natural behavior. However, since as we are using palm + mobile in this experiment, the number of possible postures will get limited to some extent and problems may arise when this system is applied to entirely different applications.

In the future experiment, it is possible that we can incorporate a higher degree of living body detection which can distinguish between a live and a fake sample with even more accuracy by combining the system used for creating the model with sensors that measure biological reaction such as ECG.

6 CONCLUSION

In this study, we investigated a palm presentation attack detection method based on an anomaly detection using Generative Adversarial Network. Our remarkable result is that the proposed PAD scheme achieved 96.8% AUC and 3% of HTER by using a model trained only with real samples. It is visible that our proposal can achieve a far better result than conventional one-class SVM systems. Additionally, it should be noted that our method can detect presentation attack by using a single static image. Therefore, this method can also be

directly applied to deal with video presentation attack or be integrated with a video-based palm liveness detection method for better performance. It is left to investigate about loss function suitable for presentation attack and reduce the number of backpropagation α to improve our method more secure and convenient.

Acknowledgement

This work was partially supported by JSPS KAKENHI Grant Number JP18K11294. We thank Eizaburo Iwata and Hiroki Kamanaka for useful discussions. The Palm Presentaton Attack Dataset is provided by Normee co, ltd.

REFERENCES

[1] A Pacut and A Czajka. "Aliveness Detection for IRIS Biometrics,". In *Carnahan Conferences Security Technology, Proceedings 2006 40th Annual IEEE International*, pages 122–129. IEEE, (2006).

[2] A. Anjos and S. Marcel. "Counter-measures to photo attacks in face recognition - A public database and a baseline,". In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–7. IEEE, (2011).

[3] M. De Marsico, M. Nappi, D. Riccio, and J. Dugelay. "Moving face spoofing detection via 3D projective invariants,". In *Biometrics (ICB), 2012 5th IAPR International Conference on*, pages 73–78. IEEE, (2012).

[4] A. Anjos, M. M. Chakka, and S. Marcel. "Motion-based counter-measures to photo attacks in face recognition,". *IET biometrics*, 3(3):147–158, (2013).

[5] H. Choi, R. Kang, and J. Choi, K.and Kim. "Aliveness Detection of Fingerprints using Multiple Static Features,". In *Proc. of World Academy of Science, Engineering and Technology*, pages 201–205, (2007).

[6] R. N. Rodrigues, N. Kamat, and V. Govindaraju. "Evaluation of biometric spoofing in a multimodal system,". In *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–5, (2010).

[7] G. Fumera G. L. Marcialis F. Roli B. Biggio, Z. Akthar. "Robustness of multi-modal biometric verification systems under realistic spoofing attacks,". In *Biometric Measurements and Systems for Security and Medical Applications (BIOMS) 2011 IEEE Workshop on*, pages 1–6. IEEE, (2011).

[8] P. Wild, P. Radu, L. Chen, and J. Ferryman. "Towards anomaly detection for increased security in multi-biometric systems: Spoofing-resistant 1-median fusion eliminating outliers,". In *IEEE International Joint Conference on Biometrics*, pages 1–6. IEEE, (2014).

[9] T. Ohki and A. Otsuka. "Theoretical vulnerabilities in map speaker adaptation,". In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2042–2046. IEEE, (2017).

[10] G. Zhang, C. Yan, X. Ji, T. Zhang, T. Zhang, and W. Xu. "DolphinAttack: Inaudible Voice Commands,". In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, CCS '17, pages 103–117. ACM, (2017).

[11] T. de Freitas Pereira, J. Komulainen, A. Anjos, J. M. De Martino, A. Hadid, M. Pietikäinen, and S. Marcel. "Face liveness detection using dynamic texture,". *EURASIP Journal on Image and Video Processing*, 2014(1):2, (2014).

[12] J. Li, T. Wang, Y.and Tan, and A. K. Jain. "Live face detection based on the analysis of fourier spectra,". In *Biometric Technology for Human Identification*, volume 5404, pages 296–304. International Society for Optics and Photonics, (2004).

[13] X. Tan, Y. Li, J. Liu, and L. Jiang. "face liveness detection from a single image with sparse low rank bilinear discriminative model,". In *European Conference on Computer Vision*, pages 504–517. Springer, (2010).

[14] G. Kim, S. Eum, J. K. Suhr, D. I. Kim, K. R. Park, and J. Kim. "Face liveness detection based on texture and frequency analyses,". In *Biometrics (ICB), 2012 5th IAPR International Conference on*, pages 67–72. IEEE, (2012).

[15] T. Ojala and D. Harwood M. Pietikainen. "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions,". In *Image Processing. Proceedings of the 12th IAPR International Conference on*, pages vol. 1, pp. 582–585. IEEE, (1994).

[16] W. R. Schwartz, A. Rocha, and H. Pedrini. "face spoofing detection through partial least squares and low-level descriptors,". In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–8. IEEE, (2011).

[17] J. Komulainen, A. Hadid, M. Pietikäinen, A. Anjos, and S. Marcel. "Complementary countermeasures for detecting scenic face spoofing attacks,". In *Biometrics (ICB), 2013 International Conference on*, pages 1–7. IEEE, (2013).

[18] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li. "A face antispoofing database with diverse attacks,". In *Biometrics (ICB), 2012 5th IAPR international conference on*, pages 26–31. IEEE, (2012).

[19] J. Määttä, A. Hadid, and M. Pietikäinen. "Face spoofing detection from single images using micro-texture analysis,". In *Biometrics (IJCB), 2011 international joint conference on*, pages 1–7. IEEE, (2011).

[20] I. Chingovska, A. Anjos, and S. Marcel. "on the effectiveness of local binary patterns in face anti-spoofing,". In *Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG-Proceedings of the International Conference of the*, pages 1–7. IEEE, (2012).

[21] G. Pan, L. Sun, Z. Wu, and S. Lao. "Eyeblink-based anti-spoofing in face recognition from a generic webcamera,". In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, (2007).

[22] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki, and A. TS Ho. "Detection of face spoofing using visual dynamics,". *IEEE transactions on information forensics and security*, 10(4):762–777, (2015).

[23] G. Pan, L. Sun, Z. Wu, and Y. Wang. "Monocular

camera-based face liveness detection by combining eyeblink and scene context,". *Telecommunication Systems*, 47(3-4):215–225, (2011).

[24] W. Bao, H. Li, N. Li, and W. Jiang. "A liveness detection method for face recognition based on optical flow field,". In *Image Analysis and Signal Processing, 2009. IASP 2009. International Conference on*, pages 233–236. IEEE, (2009).

[25] K. Lee, J. Ho, M. Yang, and K. "Visual tracking and recognition using probabilistic appearance manifolds,". *Computer Vision and Image Understanding*, 99(3):303–331, (2005).

[26] I. Chingovska and A. R. D. Anjos. "On the use of client identity information for face antispoofing,". *IEEE Transactions on Information Forensics and Security*, 10(4):787–796, (2015).

[27] J. Yang, Z. Lei, and S. Z. Yi, D.and Li. "Person-specific face antispoofing with subject domain adaptation,". *IEEE Transactions on Information Forensics and Security*, 10(4):797–809, (2015).

[28] J. I. Goodfellow, Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. "Generative adversarial nets,". In *Advances in neural information processing systems*, pages 2672–2680, (2014).

[29] A. Radford, L. Metz, and S. Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks,". *arXiv preprint arXiv:1511.06434*, (2015).

[30] T. Schlegl, P. Seeböck, Sebastian M. Waldstein, U. Schmidt-Erfurth, and G. Langs. "Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery,". In *Proceedings of the 25th International Conference of the Information Processing in Medical Imaging IPMI 2017, Boone, NC, USA*, pages 146–157, June (2017).

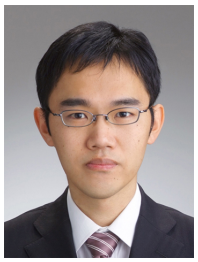
[31] K. Patel, H. Han, and A.K. Jain. "Secure Face Unlock: Spoof Detection on Smartphones,". In *IEEE Trans. Information Forensic and Security*, June (2016).

[32] A. Smola J. Shawe-Taylor J. Platt B. Schölkopf, R. Williamson. "Support vector method for novelty detection,". In *Proceedings of the 12th International Conference on Neural Information Processing Systems*, pages 582–588, (1999).

(Received October 20, 2018)
(Revised December 28, 2018)

Vishu Gupta has completed his high school from Sri Venkateshwas International School, New Delhi, India. He is currently a bachelor student within the computer science program of Faculty of Informatics, Shizuoka University. He will graduate with a B.Tech in Computer Science in 2019. His research interests include machine learning algorithms, pattern recognition and their security.

Masakatsu Nishigaki has received his Ph.D. in Engineering from Shizuoka University, Japan. He served as a Postdoctoral Research Fellow of the Japan Society for the Promotion of Science in 1995. Since 1996 he has been engaged in research at the Faculty of Informatics, Shizuoka University. He is now a Professor at the Graduate School of Science and Technology of Shizuoka University. His research interests are in wide variety of information security, especially in humanics security, media security, and network security. He served as Chief Examiner of IPSJ (Information Processing Society of Japan) Special Interest Group on Computer Security from 2013 to 2014, Chair of IEICE (Institute of Electronics, Information and Communication Engineers) Technical Committee on Biometrics from 2015 to 2016, and currently serving as Director of JSSM (Japan Society of Security Management) since 2016. He is IPSJ (Information Processing Society of Japan) fellow.



Tetsushi Ohki has received the BE and ME degrees in electronics and communication engineering from Waseda University, Tokyo, Japan, in 2002 and 2004, respectively, and the Ph.D. degree in Engineering from Waseda University in 2010. He is currently a Lecturer at the Graduate School of Science and Technology of Shizuoka University, Japan. His research interests include biometrics, pattern recognition, information security and privacy. He is a member of IEICE (Institute of Electronics, Information and Communication Engineers) and IPSJ (Information Processing Society of Japan) .

Regular paper

A Proposal of a Method for Video Advertisement Insertion on Smartphone

Yoshia Saito*

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

Abstract –
In this research, we propose a method for video advertisement insertion on smartphone. The method has an algorithm which estimates a comfort timing to insert a video advertisement using the acceleration data at the time of watching a video. To create the algorithm, we formulated three hypotheses; (1) Viewers who watch video contents on smartphones sitting on chairs change their posture when they take a short breath, (2) The postural change can be detected by analyzing acceleration data from the accelerometer of the smartphone, (3) The timings of the postural changes correspondent with timings of a scene change on the video content and one of the timings is an appropriate timing to insert a video advertisement. We conducted a preliminary experiment and found these hypotheses were true. On the basis of the finding, we proposed an algorithm which estimates a comfort timing using the acceleration data to detect a viewer's postural change. The evaluation results showed accuracy rate of the proposed algorithm was 86% and useful in terms of practical usage.

Keywords: Video Advertisement, Insertion Algorithm, Smartphone, Accelerometer

1 INTRODUCTION

In recent years, video sharing services introduce a business model which inserts video advertisements in their video contents in the same manner as TV. The business model becomes popular with increase of smartphone users. Major video sharing services such as YouTube [1] and niconico [2] provide smartphone applications to users and the smartphone applications insert video advertisements to make a profit. Therefore, many smartphone users have to watch video advertisements on the video sharing services.

There are three types of video advertisements, which are pre-roll, post-roll and mid-roll. The pre-roll video advertisement inserts a video advertisement before video start. The post-roll inserts a video advertisement after video end. The mid-roll inserts a video advertisement viewing a video content like TV commercials and becomes popular in recent years. Adobe reports the mid-roll video advertisement is engaging commercials which have high completion rate [3]. The mid-roll video advertisements will be used even further in the next several years.

The mid-roll video advertisement, however, creates disadvantage for viewers. If video advertisements are inserted at the wrong time, the viewers feel discomfort about the advertisement and it will reduce effectiveness of the advertisement. Furthermore, the discomfort about the advertisement gives the viewers cause to use adblock software. There are over 600 million devices running adblock software and 11% of the global internet population is blocking ads on the web [4]. Moreover, Google Chrome starts to block web advertisements which do not conform to Better Ads Standards defined by the Coalition for Better Ads [5]. The advertisements ads which disrupt users' experience tend to be excluded from web services. Therefore, it is important to insert video advertisements at right timing so that the viewers can watch video contents comfortably and the advertisements do not disrupt their experiment. Otherwise, the video advertisements will be blocked and meaningless.

We have proposed a video advertisement insertion method which does not interfere with video viewing to make viewers accept the video advertisements [6]. In the previous work, it analyzes characteristics of viewers' comments for the video content. It enables viewers to watch videos more comfortably without feeling of interruption of their video viewing by the video advertisement insertion. However, there are two issues in the previous method. The first issue is that the previous method does not apply to various video sharing services. This is because it needs special kind of viewers' comments with playback time such as comments on the niconico. The second issue is that the previous method has room for improvement of viewers' experience. This is because it does not personalize the timing of video advertisement insertion in spite of difference of the right timings for each viewer.

In order to solve these issues, based on the fact that the number of users watching videos on smartphones has increased rapidly, it is worthwhile considering a method using smartphone sensor information. In this research, we use acceleration information from the accelerometer of the smartphone at the time of watching video contents. We try to find relationship between the acceleration information and appropriate timings for video advertisement insertion. Utilizing the relationship, we propose a method which estimates an appropriate timing to insert a video advertisement for each viewer.

The paper is organized as follows. In the next section, we describe related work and previous work to clarify the issues. In section 3, we formulate hypotheses in order to create a new method which utilizes the accelerometer of the smartphone. Then, we conduct a preliminary experiment in section 4. We propose a new video advertisement insertion method in section 5. In section 6, we evaluate the new method compared with the previous method. Section 7 gives some conclusions and our future work.

2 RELATED WORK

In this section, we describe an advertising model which becomes a reason why the viewers' discomfort should be removed. Then, we describe related work of video advertisement insertion and detail of our previous work.

2.1 Advertising Model

There is an advertising model, AIDMA [7] which is a psychological process model that leads consumers to purchase some products. The process goes along "Attention", "Interest", "Desire", "Memory" and "Action". At first, consumers watch advertisements and aware of a products (Attention). Then, they are interested in it (Interest) and desire to get it (Desire). They memorize the product (Memory) and purchase it at last (Action).

In viewpoint of video sharing services, wrong timing of video advertisement insertion has a negative impact on the process. If the viewers feel uncomfortable about the video advertisement which is inserted at the wrong time, the process will stop at "Attention" phase and not proceed to "Interest" phase. Therefore, it is important to insert video advertisements at right timing.

2.2 Video Advertisement Insertion

There are studies of interactive advertising to provide interactivity to the advertising [8]-[11]. The interactive advertising allows selecting appropriate ads according to the viewers and changing video length and display methods. Our research is regarded as one of technologies for interactive advertising. Tao Mei et al. [12] proposed a scheme of appropriate video ad insertion for online videos. In this research, the appropriate timing for video advertisement insertion is determined detecting an unattractive video shot boundary. The unattractive video shot boundary is detected by importance of the scene audio-visually. Since this research analyzes video image and audio in detail, the processing cost will be high when it applies to a large number of videos on the video sharing services. Our study aims to find other approaches which estimate the appropriate timing for video advertisement insertion without heavy audio-visual processing.

2.3 Our Previous Work

We have proposed a method for video advertisement insertion which was applicable to the action game videos using viewers' comments for the video contents on the niconico [6]. We found appropriate timings for video advertisement insertion corresponded with timings of scene changes on the video content. The timings of the scene changes could be estimated by analyzing viewers' comments. Details of the method is as follows.

1. Getting 10,000 viewers' comments of the video.
2. Detecting shot boundaries of the video.
3. Calculating sample variance of number of viewers' comments per second from first shot boundary to last shot boundary.
4. Omitting the first shot and the last shot from the following process.
5. Detecting a shot boundary which is first with lowest sample variance after the first boundary with maximum sample variance from start of the video.
6. Referring 7 shots centered at the detected shot boundary.
7. Calculating variance of number of viewers' comments per second for each shot.
8. Calculating difference of the variance between two adjacent shots.
9. Finding maximum difference of the variance.
10. Choosing the shot boundary for video advertisement insertion.

However, the previous method does not apply to various video sharing services because it needs special kind of viewers' comments with playback time such as comments on the niconico. Moreover, it does not personalize the timing of video advertisement insertion in spite of differences of right timing for each viewer. In this paper, we try to solve these two issues and propose a method which does not need viewers' comments with an appropriate timing of video advertisement insertion for each individual.

3 HYPOTHESES FORMULATION

In this section, we indicate possibility that there is a relationship between human motion and degree of interest. We also describe existing techniques of sensing for human motion to choose what sensor is appropriate for estimating human motion on smartphones. Then, we formulate hypotheses in order to create a new method.

3.1 Human Motion and Degree of Interest

There are several techniques for estimating human motion by sensor devices on the smartphone. There are also some studies which show a relationship between human motion and degree of interest.

The relationship between eye motion and degree of interest is well-known [14]-[17]. The data of eye motion can be acquired by eye-tracking techniques. However, high accurate eye-tracking techniques require special devices for eye tracking or strict restrictions of the measuring environment. It is difficult for smartphone users to prepare for the special devices and strict restrictions force inconvenience upon the users. For these reason, the eye motion is not suitable to estimate viewers' degree of interest on the smartphone.

The relationship between posture and degree of interest is mentioned [18]. People change their posture at intervals from 15 to 20 minutes at the time of sitting because of fatigue [19][20]. Meanwhile, their postural change hardly occurs when they are interested on something. We can apply this knowledge to a method which estimates an appropriate timing to insert a video advertisement for each viewer if we can detect viewers' postural change by smartphone sensors.

There are a lot of techniques to estimate body motion using sensors. Visual analysis using video data taken by camera devices is one of the techniques. However, usage of camera devices causes large power consumption and it is a disadvantage especially on the smartphone. Visual analysis is not appropriate to estimate body motion. Estimation of the body motion using acceleration information is popular and lightweight techniques [21][22]. Most of smartphones have accelerometer and many researchers study estimating body motion of the smartphone users from the acceleration information. These researches show various states of smartphone users such as sitting, standing, walking, running, going up and down the stairs and so on can be discriminated. Usage of accelerometer to detect postural change of smartphone users is reasonable.

3.2 Hypotheses

Our previous work shows appropriate timings for video advertisement insertion corresponded with timings of scene changes on the video content. The viewers may change their posture in scene changes on the video content. We formulate three hypotheses as follows.

1. Viewers who watch video contents on smartphones sitting on chairs change their posture when they take a short breath.
2. The postural change can be detected by analyzing acceleration data from the accelerometer of the smartphone.
3. The timings of the postural changes correspondent with timings of a scene change on the video content and one of the timings is an appropriate timing to insert a video advertisement.

If these hypotheses are true, we can create a new method which estimates an appropriate timing for each individual to insert a video advertisement by analyzing acceleration data from the accelerometer of the smartphone.

4 PRELIMINARY EXPERIMENT

We conduct a preliminary experiment to reconfirm if there are difference of right timings to insert a video advertisement for each viewer and test the hypotheses.

4.1 Methodology

We select 5 videos at random from several videos which were used in the previous work. We ask 3 participants for cooperation, who are in their twenties and thirties and have used some video sharing services. At first, we explain about the experiment to the participants. They watch the 5 videos in random order on a smartphone sitting on a chair. We shoot a video to record their watching situation. In the smartphone, an application which records acceleration data from the accelerometer is running. After watching the videos, we carry out a questionnaire survey to ask the participants top 3 comfort timings if a video advertisement is inserted. We explain the participants the length of the video advertisement is about 15 seconds. Then, we interview the participants showing the recorded video. Table 1 shows the overview of the experiment.

Table 1: Overview of the experiment.

Participants	3 people in 20s and 30s who have used video sharing services
Smartphone	Nexus 5
Videos	Video 1: 3D action game [23] Video 2: 2D action game [24] Video 3: 3D action game [25] Video 4: 2D action game [26] Video 5: 3D action game [27]
Condition	Sitting on a rotary chair with backrest (seat height: 30 cm)
Procedure	1. Receive an explanation about the overview of the experiment 2. Watch the 5 videos in random order 3. Answer top 3 comfort timings if a video advertisement is inserted 4. Take an interview about the comfort timings.

Table 2: Comfort timings for each participant in video 1.

Video 1

	Participant A	Participant B	Participant C
1st comfort timing	01:47	07:00	01:08
2nd comfort timing	00:09	06:43	01:47
3rd comfort timing	06:30	01:47	06:43

Table 3: Comfort timings for each participant in video 2.

Video 2

	Participant A	Participant B	Participant C
1st comfort timing	01:00	09:49	01:23
2nd comfort timing	01:41	10:41	10:41
3rd comfort timing	09:41	05:45	09:41

4.2 Results

At first, we reconfirm whether there are differences of right timings to insert a video advertisement for each viewer or not. Tables 2-6 show the top 3 comfort timings in the 5 videos for each participant. These results show there are differences of the right timing for each participant. We found the participants had characteristic features to select the comfort timings. The participant A tended to select the 1st comfort timing in early scenes. The participant B tended to select the 1st comfort timing in late scenes. We reconfirmed that it was necessary to personalize the timing of video advertisement insertion because there were difference of the right timings for each viewer.

We also checked the recorded video. The participants changed their posture in the scene change and the timings of the postural change matched their comfort timing. The postural changes rapidly increased the resultant acceleration of 3-axis. These results show monitoring the rapid increase of the resultant acceleration can detect postural changes of the

Table 4: Comfort timings for each participant in video 3.

Video 3

	Participant A	Participant B	Participant C
1st comfort timing	01:23	10:50	07:09
2nd comfort timing	02:19	11:11	10:50
3rd comfort timing	07:09	06:14	11:11

Table 5: Comfort timings for each participant in video 4.

Video 4

	Participant A	Participant B	Participant C
1st comfort timing	01:39	07:10	01:39
2nd comfort timing	01:44	09:38	07:10
3rd comfort timing	07:10	10:16	03:45

Table 6: Comfort timings for each participant in video 5.

Video 5

	Participant A	Participant B	Participant C
1st comfort timing	03:41	10:51	05:44
2nd comfort timing	05:44	06:50	06:50
3rd comfort timing	06:50	11:15	11:15

viewers and estimate one of their comfort timings to insert a video advertisement in the video content.

5 PROPOSED METHOD

The preliminary experiment shows the possibility of estimating a comfort timing for each viewer to insert a video advertisement utilizing their postural change which can be detected by the rapid increase of the resultant acceleration of 3-axis. On the basis of the finding, we create an algorithm for video advertisement insertion.

5.1 Algorithm for Video Advertisement Insertion

Figure 1 shows a flowchart of the algorithm for video advertisement insertion using an accelerometer on the smartphone. In the algorithm, it calculates a test statistic based on acceleration values for outlier detection which means occurring the viewer's postural change. The test static T_i can be calculated using $Acceleration_t$, E_t, SD_t at time t. $Acceleration_t$ denotes the resultant acceleration of 3-axis at time t. E_t denotes the average of the resultant acceleration from the video start to time t. SD_t denotes the standard deviation of the resultant acceleration from the video start to time t. T_t is calculated by the following equation.

$$T_t = | (Acceleration_t - E_t) | / SD_t$$

The outlier can be detected when the following inequality is completed.

$$T_t > 2 * SD_t$$

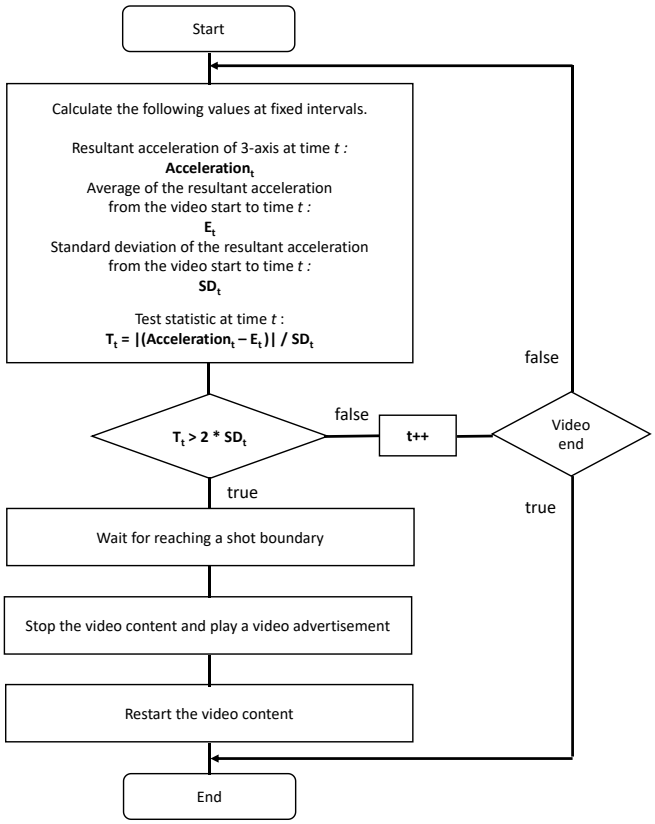


Figure 1: The flowchart of the algorithm for video advertisement insertion.

If the outlier is detected, the algorithm waits for reaching a next shot boundary. In the next shot boundary, the video content is stopped and a video advertisement starts. After completion of the video advertisement, the video content restarts and the algorithm terminates the process of the video advertisement insertion.

5.2 System Design

We assume the algorithm for video advertisement insertion is implemented on the smartphone as an application for video viewing. A video advertisement and acceleration data are input to the algorithm. The algorithm outputs less than one comfort timing for each viewer in the video content.

Figure 2 shows the system design of the proposed method. The proposed system have two servers, which are for video sharing and video advertisement. Video uploaders submit their video contents to the video sharing server. Advertising sponsors provide video advertisements to the video advertisement server. Video viewers have smartphones with a smartphone application for video viewing which has the algorithm for video advertisement insertion. The smartphone application plays a video content from the video sharing server. Shot boundaries of the video content are detected by existing techniques for shot boundaries detection and it lies outside the scope of our research. While the video viewer is watching the video content, the algorithm for video advertisement insertion

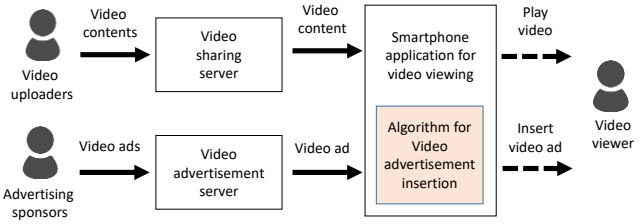


Figure 2: System design of the proposed method.

monitors acceleration data of the smartphone. The smartphone application stops the video content temporarily and starts a video advertisement from the video advertisement server when the algorithm estimates the timing is comfort for the viewer. After that, the video content restarts without any more video advertisement in the viewing. Note that we suppose the algorithm works only when the video viewer is sitting on a chair and the viewer watches the video holding the smartphone without video skip. In case of other conditions, new routines should be added to the algorithm.

6 EVALUATION

We evaluate the performance of the proposed algorithm for video advertisement insertion using an accelerometer on the smartphone. Comparing the previous algorithm, we verify the proposed algorithm estimates a better comfort timing for each viewer.

6.1 Methodology

The evaluation is conducted in the same manner as the preliminary experiment. We use 5 videos which are same videos in the preliminary experiment. We ask 10 participants for cooperation, who are in their twenties and thirties and have used some video sharing services. At first, we explain about the experiment to the participants. They watch the 5 videos in random order on a smartphone sitting on a chair. We shoot a video to record their watching situation. In the smartphone, an application which records acceleration data from the accelerometer is running. After watching the videos, we carry out a questionnaire survey to ask the participants top 3 comfort timings if a video advertisement is inserted. We explain the participants the length of the video advertisement is about 15 seconds.

After getting data of acceleration and comfort timings, we estimated a comfort timing for each viewer by using the proposed algorithm. We also estimated a comfort timing by using the previous algorithm which used viewers' comments on the niconico. Then, we compared the result of the proposed algorithm with one of the previous algorithm.

6.2 Results

Figure 3 shows graphs of the resultant acceleration of 3-axis on the smartphone when each participant was viewing the Videos 1-5. The horizontal axis and vertical axis of these graphs show elapsed time since the video started and the resultant acceleration of 3-axis. The proposed algorithm for video advertisement insertion estimated a comfort timing for each viewer based on the acceleration data. From these graphs, we can see the rapid increase of the resultant acceleration caused by the participant's postural change.

Table 7 shows the result of comparing an estimated timing of the proposed algorithm with top 3 comfort timings for each viewer. Table 8 shows the result of comparing an estimated timing of the previous algorithm with top 3 comfort timings for each viewer. In these tables, "1st" means the estimated timing coincides with the 1st comfort timing for the viewer. The same applies to "2nd" and "3rd". "n/a" means the estimated timing does not coincide with any top 3 comfort timings. We regard the estimated timing is an appropriate timing if it coincides with one of the top 3 comfort timings.

From Table 7, the proposed algorithm could estimate 43 appropriate timings of the 50 chances. The accuracy rate of the proposed algorithm was 86%. On the other hand, the previous algorithm could estimate only 22 appropriate timings of the 50 chances as shown in Table 8. The accuracy rate of the previous algorithm was 44%. Comparing these results, the proposed algorithm improved the accuracy rate more than 40%. Personalization of a timing to insert a video advertisement contributed the improvement of accuracy rate.

Table 7: Result of comparing an estimated timing of the proposed algorithm with top 3 comfort timings for each. viewer

	Video 1	Video 2	Video 3	Video 4	Video 5
Participant A	1st	2nd	1st	2nd	n/a
Participant B	1st	1st	2nd	n/a	n/a
Participant C	1st	3rd	1st	1st	1st
Participant D	3rd	1st	3rd	n/a	1st
Participant E	1st	1st	2nd	1st	3rd
Participant F	1st	1st	1st	2nd	2nd
Participant G	2nd	3rd	1st	1st	n/a
Participant H	1st	2nd	1st	n/a	1st
Participant I	1st	n/a	3rd	2nd	3rd
Participant J	2nd	1st	1st	1st	1st

Table 8: Result of comparing an estimated timing of the previous algorithm with top 3 comfort timings for each. viewer

	Video 1	Video 2	Video 3	Video 4	Video 5
Participant A	n/a	n/a	n/a	1st	n/a
Participant B	n/a	n/a	1st	n/a	3rd
Participant C	n/a	n/a	2nd	1st	n/a
Participant D	n/a	n/a	3rd	3rd	n/a
Participant E	n/a	3rd	1st	3rd	n/a
Participant F	n/a	2nd	3rd	n/a	2nd
Participant G	3rd	n/a	1st	2nd	n/a
Participant H	n/a	n/a	1st	n/a	2nd
Participant I	n/a	n/a	n/a	3rd	n/a
Participant J	3rd	3rd	1st	n/a	n/a

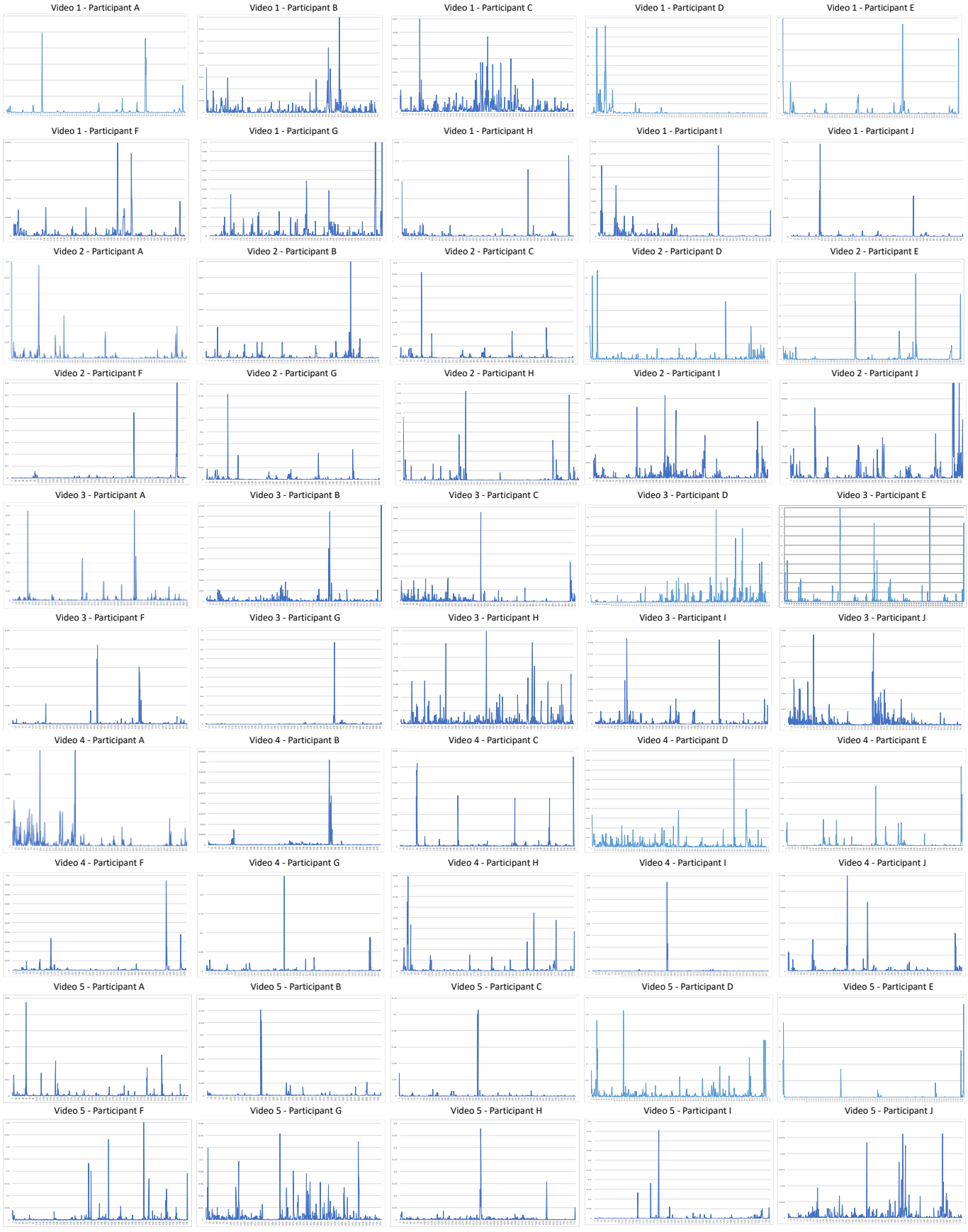


Figure 3: The resultant acceleration of 3-axis in Videos 1-5.

Table 9: Result of top 3 videos which the participants answered to hesitate to be watch a video advertisement.

	1st negative opinion of ad insertion	2nd negative opinion of ad insertion	3rd negative opinion of ad insertion
Participant A	Video 5	Video 4	Video 2
Participant B	Video 5	Video 4	Video 3
Participant D	Video 4	Video 1	Video 3
Participant E	Video 5	Video 3	Video 4
Participant G	Video 2	Video 5	Video 4
Participant H	Video 4	Video 2	Video 5
Participant I	Video 4	Video 5	Video 3

We interviewed the participants about the results. *Participant I* answered the estimated timing of the proposed algorithm in Video 2 was 7th comfort timing. This is not so comfort but an acceptable timing. From this fact, all participants will accept all estimated timings of the proposed algorithm in Videos 1-3. However, there are many unacceptable estimated timings of the proposed algorithm in Videos 4-5. To study a reason why the proposed algorithm did not estimate comfort timings, we also interviewed the participants if there were videos which they would like not to watch a video advertisement. 7 participants answered "Yes". Then, we ask top 3 videos in which they hesitate to be watch a video advertisement. Table 9 shows the result of top 3 videos which the participants answered to hesitate to be watch a video advertisement. From the result, most of the participants answered Videos 4-5 are not suitable for video advertisement insertion. The reason why they does not like to be inserted a video advertisement in Video 4-5 is "There is not clear scene changes in these videos". When there were not clear scene changes in a video content, it was better not to use the mid-roll video advertisement in the first place and the low accuracy rate of the proposed algorithm was unavoidable result.

7 CONCLUSION

In this paper, we proposed a method for video advertisement insertion using acceleration data on smartphone. From the preliminary experiment, we got 3 findings; (1) viewers who watch video contents on smartphones sitting on chairs change their posture when they take a short breath, (2) the postural change can be detected by analyzing acceleration data from the accelerometer of the smartphone, (3) the timings of the postural changes correspondent with timings of a scene change on the video content and one of the timings is an appropriate timing to insert a video advertisement. Then, we created an algorithm for video advertisement insertion and designed its system.

From the evaluation results, we found the proposed algorithm improved the accuracy rate more than 40% comparing with the previous algorithm. This result showed the effectiveness of the proposed algorithm. However, we also found the proposed algorithm could not estimate an appropriate comfort timing in the videos which did not have clear scene changes because it was inappropriate to use the mid-roll video advertisement.

For the future work, we will try to study other algorithms even if the viewer does not sit on a chair. We also study a method which switch mid-roll to pre-roll or post-roll video advertisement when there are not clear scene changes in the video content.

REFERENCES

[1] YouTube, http://www.youtube.com/
[2] niconico, http://www.nicovideo.jp/
[3] 2012 adobe digital video advertising report, https://blogs.adobe.com/primetime/files/2013/11/Monetization-Report_FINAL1.pdf
[4] 2017 Adblock Report, https://pagefair.com/blog/2017/adblockreport/
[5] Coalition for Better Ads, https://www.betterads.org/
[6] Y. Saito, "A method for video advertisement insertion with audience comments on action game videos", International Workshop on Informatics (IWIN2017), pp.153-158 (2017).
[7] S. Roland Hall, "Retail advertising and selling", The History of advertising: 40 major books in facsimile (1985).
[8] K. Risden, M. Czerwinski, S. Worley, L. Hamilton, J. Kubiniec, H. Hoffman, N. Mickel and E. Loftus, "Interactive advertising: patterns of use and effectiveness", SIGCHI, pp. 219-224 (1998).
[9] J. W. Kim and S. Du, "Design for an Interactive Television Advertising System, Proceedings of the 39th Annual Hawaii International Conference on System Sciences", Vol. 2 (2006).
[10] J. Lloyd, "I-Ads - a new approach, European Conference on Interactive Television" (2003).
[11] P. Giotis, G. Lekakos, "Effectiveness of Interactive Advertising Presentation Models", EuroITV '09, pp.157-160 (2009).
[12] T. Mei, "VideoSense-Towards Effective Online Video Advertising", ACM Multimedia'07, pp.1075-1084 (2007).
[14] J. Heer and S. K. Card, "Efficient user interest estimation in fisheye views", CHI '03 Extended Abstracts on Human Factors in Computing Systems, pp. 836-837 (2003).
[15] A. Santella and D. DeCarlo, "Robust clustering of eye movement recordings for quantification of visual interest", Proceedings of the 2004 symposium on Eye tracking research & applications, pp. 27-34 (2004).
[16] T. Walber, C. Neuhaus, S. Staab, A. Scherp and R. Jain, "Creation of individual photo selections: read preferences from the users' eyes", Proceedings of the 21st ACM international conference on Multimedia, pp. 629-632 (2013).
[17] V. Georges, F. Courtemanche, S. Senecal, T. Baccino, M. Fredette and P. Leger, "UX Heatmaps: Mapping User Experience on Visual Interfaces", Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 4850-4860 (2016).

[18] S. Mota and R. W. Picard, “Automated Posture Analysis for detecting Learner’s Interest Level”, Computer Vision and Pattern Recognition Workshop (2003)
[19] P. Branton, “Behaviour, Body mechanics and Discomfort, Ergonomics,” Vol.12, No.2 (1969).
[20] H. Watanabe, M. Ando and T. Takahashi, “Transition of sitting posture over time”, J. Archit. Plann. Environ. Eng., AIJ, No. 474, pp. 107-114 (1995).
[21] T. Iso and K. Yamazaki, “Gait analyzer based on a cell phone with a single three-axis accelerometer”, ACM the 8th Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI2006), pp. 141-144 (2006).
[22] R. Slyper and J. K. Hodgins, “Action capture with accelerometers”, Proceedings of the 2008 ACM SIGGRAPH/Eurographics, pp. 193-199 (2008).
[23] niconico, http://www.nicovideo.jp/watch/sm5457137
[24] niconico, http://www.nicovideo.jp/watch/sm6979644
[25] niconico, http://www.nicovideo.jp/watch/sm2750853
[26] niconico, http://www.nicovideo.jp/watch/sm4895582
[27] niconico, http://www.nicovideo.jp/watch/sm8481759

(Received October 21, 2018)

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

Submission Guidance

About IJIS

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

Aims and Scope of Informatics Society

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

Areas of interest include, but are not limited to:

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

Instruction to Authors

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

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

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

LaTeX2e

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

Microsoft Word™

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

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

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

Copyright

Publisher

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

E-mail: secretariat@infsoc.org

CONTENTS