International Journal of Informatics Society

01/26 Vol.17 No.2 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 quality and value as a resource. Informatics also referred to as Information science, studies the structure, algorithms, behavior, and interactions of natural and artificial 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 the study of informatics 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

Yuichi Tokunaga

Guest Editor of the Fifties Issue of the International Journal of Informatics Society

We are delighted to have the Fifties issue of the International Journal of Informatics Society (IJIS) published. This issue includes selected papers from the Eighteenth International Workshop on Informatics (IWIN2024), which was held at Yanagawa, Fukuoka, September 1st – 4th, 2024. The workshop was the eighteenth event for the Informatics Society. It 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 six technical sessions. The workshop was successfully finished, and precious experiences were provided to the participants. It highlighted the latest research results in informatics and its applications, including networking, mobile ubiquitous systems, data analytics, business and industrial systems, education systems, design methodology, intelligent systems, groupware, and social systems, etc.

Each paper submitted to IWIN2024 was reviewed in terms of technical content, scientific rigor, novelty, originality, and presentation quality by at least two reviewers. Through those reviews, 22 papers were selected for publication candidates of the IJIS Journal, and they were further reviewed as Journal papers. We have three categories of IJIS papers, Regular papers, Practical papers, and Invited papers, each of which was reviewed from different points of view. This volume includes 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 all over the world.

Yuichi Tokunaga received his Ph.D. in science and engineering from Ritsumeikan University in 2009. He joined Mitsubishi Electric Corporation in 1990 and engaged in the R&D of high-reliability computers, wireless sensor networks, the algorithms of time-synchronization and positioning, the network protocol for industrial applications, and data analytics for condition-based maintenance. He has been a professor at Kanazawa Institute of Technology since 2019. He is a member of the ITS steering committee of IPSJ and a member of IEEE and ISCIE.

Regular Paper

Aerial Photography Planning Method for 3D Model Creation to Realize Remote Inspection for Utility Poles

Koji Yamagishi*, Yuichi Tokunaga*

*Graduate School of Engineering, Kanazawa Institute of Technology, Japan
{c6301756@st, y.tokunaga@neptune}.kanazawa-it.ac.jp

Abstract - In this paper, we propose and verify a method for efficiently generating flight paths for aerial photography using drones, with the aim of reducing the labor required for inspecting utility poles in Japan. Utility poles are critical infrastructure supporting power and telecommunications networks; however, due to their large number and wide distribution, they are currently inspected manually. To carry out 3D model inspections for utility poles using drones, which are used for bridges and dams, it is necessary to plan the flight path and aerial photography points so that many utility poles can be photographed efficiently in a single flight.

Therefore, we devised an algorithm that aims to pass through the theoretically optimal photography position of the utility pole and searches for the optimal photography point on the main flight route while prioritizing constant-speed flight, thereby improving the model generation accuracy. Through computer evaluation, we confirmed that, compared to previous studies that assumed multirotor drones, the model accuracy of VTOL drones and this algorithm is equivalent, but flight time can be reduced by 72.6%.

Keywords: 3D model, Drone, Utility pole, Structure from Motion, Maintenance

1 INTRODUCTION

In recent years, several social infrastructure facilities have waged over 50 years since their construction, and the useful life of equipment is also considered to be 50 years [1].

For example, the United States has been developing infrastructure on a large scale since the New Deal policies in the 1930s [2]. Fifty years later, in the 1980s, the U.S. faced the fundamental problem of aging infrastructure that impacted various aspects of the economy and daily life [3]. Japan's social infrastructure was constructed intensively during rapid economic growth between 1955 and 1970. Fifty years later, Japan already has many aging facilities, and this number is expected to increase at an accelerating rate [4].

Maintenance of these facilities, including inspection and repair, is essential. Inspection is the task of evaluating the condition of equipment, and with inspection comes the process of selecting equipment to be repaired. The efficiency and accuracy of this process has a significant impact on overall maintenance. Several efficient inspection methods have been developed for use in various facilities. However, few studies have focused on inspection methods for utility poles, which are small structures. Utility poles are more numerous in larger facilities. The total number of utility poles in 2016 was 35.78 million and increased by approximately 70,000 per year. Of these, 67% are electric poles for power transmission and 33% are telegraph poles for telecommunications. Time, cost, and human efficiency must be emphasized when inspecting all utility poles.

Conventionally, utility pole inspections have relied on visual inspection characterized by high reliability and accuracy. However, visual inspections require inspectors to visit all utility poles on site, which is challenging because geographical factors and the height of the poles make it difficult to assess their condition. Consequently, visually inspecting all utility poles is inefficient in terms of time, money, and workforce.

Inspection methods that use three-dimensional (3D) models are being considered for large facilities. Tion et al. described a 3D reconstruction method for multi-vision-based inspection of pipelines [5]. They also presented an efficient and accurate 3D reconstruction method for industrial pipeline inspections using photogrammetry. The five photogrammetric methods used in this study were stereovision, photometric stereo, CAD-based photogrammetry, motion-based photogrammetry, and shading-based photogrammetry.

In addition, drones are used to acquire data for 3D model generation. Miyake et al. used a drone to capture aerial shots of a road's slope, fill, and a bridge to create a 3D model representing structures, slopes, and fills on the road [6]. This allowed identifying signs based on the differences between the current and past 3D models.

There are other reasons why labor-saving needs to be achieved. The number of construction workers, including infrastructure inspectors, has been decreasing since 1997, with 5 million in 2015, which is 73% of the peak; this shortage is expected to continue [7]. Therefore, as with large facilities, drones and 3D models are expected to be utilized; however, as mentioned earlier, utility poles need to be made more efficient in terms of their number.

Therefore, this study examines a method for generating high-precision 3D models of utility poles using aerial images captured by drones to improve the efficiency. Creating a 3D model of an entire utility pole using aerial photography with a drone allows inspections to be conducted remotely. As a result, inspections without relying on manpower become possible, and the efficiency improves.

On the other hand, the use of VTOL drones is being considered for wide-area inspections [8]. A VTOL drone can take off and land vertically like a helicopter and can cruise at high speeds, such as a fixed-wing aircraft. It can take off and land in confined spaces and is characterized by its ability to

travel long distances at high speeds [9]. Compared to a normal multirotor drone, a VTOL drone cannot perform maneuvers, such as hovering or turning. However, because they are able to travel long distances at high speeds, they can inspect a wide area within a short time.

Therefore, in this study, we propose an aerial photography point and a movement path generation algorithm suitable for VTOL drones. The contributions to this paper are summarized as follows:

- VTOL drones have the advantage of being more efficient with their batteries, as they have a longer range than multirotor drones; however, they also have the drawback of not being able to make sharp turns. By overcoming this disadvantage with the proposed algorithm, we were able to gain prospects for using VTOL drones to inspect the utility poles.

Section 2 introduces 3D models and related studies on their use and generation using aerial photography. Section 3 describes the proposed framework based on the hypotheses derived from these studies, and Section 4 describes the evaluation. Finally, conclusions are presented in Section 5.

2 RELATED WORK

In this section, previous drone 3D model generation methods are presented, and the requirements for 3D modeling of utility poles are clarified.

A 3D model comprises point-cloud data obtained using laser scanning or photogrammetry. Laser scanning is generally performed using LiDAR and 3D laser scanners. The time-of-flight method, which measures the time of travel due to laser emission and reflection, is used to obtain the spatial coordinates of an object's surface. Therefore, spatial coordinates can be accurately measured within an error range of 1–10 mm [10]. However, laser scanners are expensive, and their use may be limited by certain circumstances that can distort measurements, such as laser penetration and diffuse reflection.

The photogrammetric method acquires point-cloud data by reconstructing 2D images with overlapping intervals into a 3D point cloud. Point-cloud data are acquired using techniques such as Structure from Motion (SfM) and Multiview Stereo (MVS). SfM obtains spatial coordinates by superimposing multiple photos and calculating distances based on critical points in the photos. MVS refers to the recovery of a 3D shape from a camera's captured point. Photogrammetry methods are challenging because of their low accuracy.

However, photogrammetric methods are attracting attention as alternatives to LiDAR because of their low cost and ease of acquiring point clouds [11]. Therefore, various mobility-based imaging methods are being considered to efficiently acquire data in terms of time, money, and manpower.

Drones and other uncrewed aircraft are used for data acquisition. Yoon et al. proposed an unmanned aerial vehicle-based missing area detection and damage location method based on 3D image coordinates as an alternative to the visual inspection of piers [12]. Sungsik's methodology comprises the following three phases. In Phase 1, coordinate transformation is performed using the point and attitude information of the uncrewed aircraft and camera, and the distance information between the camera and the target surface is used to obtain the coordinates of the center point of each acquired image. In Phase 2, the focal length and working distance of the camera were used to calculate the field-of-view size for each acquired image. In Phase 3, information regarding the size of the field of view of all images computed in the previous phases was used to identify missing portions of the region of interest. This makes it possible to locate the damage detected at the individual image level in the entire inspection area.

The results showed that the missing areas and damaged locations could be identified with an error of 10 cm while leveraging the efficiency of uncrewed aerial vehicles. However, because utility poles are small-scale facilities, the 10-cm error becomes significant and cannot be ignored. Eliminating this error is necessary to generate a high-quality 3D model.

Moritani et al. proposed a method for estimating camera points for additional photography based on the results of low-quality region extraction as a guideline for optimal photography planning for efficient and high-quality 3D model generation using SfM-MVS [13]. This study focused on the fact that it is currently necessary to determine in advance the number of captured images from a location to generate a high-quality 3D model. In addition, if too many images are captured, the MVS process requires a significant amount of time. A related study used the procedure shown in Fig. 1 to predict the quality of a high-density model generated by the MVS process. The prediction relies solely on 3D tie points and camera poses estimated by the SfM process, which can be computed quickly. The study also estimated the optimal additional image capture points to improve the quality of regions predicted to be of low quality. The goal is to estimate the optimal additional image-capturing locations that can improve the quality of the predicted regions.

The specific procedure is to calculate additional candidate photography points where orthographic imaging of the low-quality area is possible from the center of gravity point of the low-quality area and then obtain an index of candidate

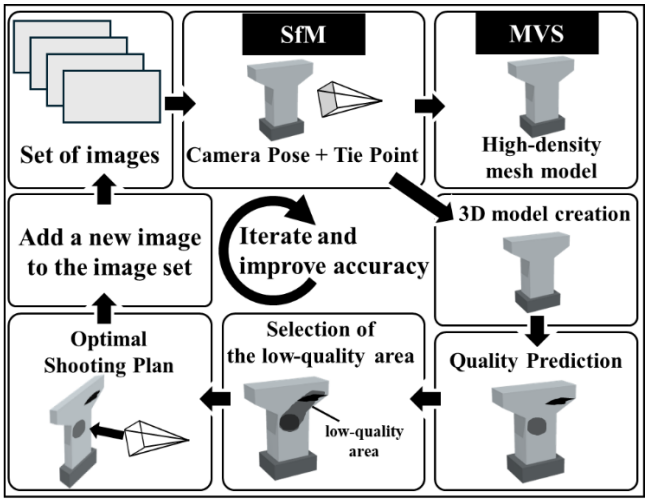


Figure 1: SfM-MVS process and proposed process.

additional photography points. The relationship between the calculated candidate additional photography points and the photography points that best covered the low-quality area among the already captured cameras was expressed as a particular index. The two indices were weighed and combined. This procedure was performed for all the candidates' additional photography points, and the additional photography points corresponding to the index with the lowest value were selected as the optimal point. Repeating this method allows for the efficient determination of additional photography points.

We compared the accuracy of the 3D model with three groups of images: 30 randomly selected images, 72 randomly selected images, and 72 images selected using the proposed method, from a set of 171 original images captured densely around the outer circumference of the piers. The average error of the 30 randomly selected images was 12.3 mm, and that of the 72 selected images was 6.1 mm. By contrast, the proposed method with 72 images resulted in an error of 2.2 mm, indicating its effectiveness method for 3D model accuracy. However, although high accuracy can be obtained, the number of shots is large, and should be reduced while maintaining the accuracy of the 3D model.

Yamazaki et al. proposed an optimal acquisition planning method for 3D reconstruction using SfM/MVS for construction sites [14]. This study used the "model-based methods" method, a simple 3D model based on geospatial information to plan optimal photography locations and routes [15]. The specific process was to homogenize the simplified model by meshing. Next, from the candidate voxels for photography points, points that minimized the number of shots required for modeling were extracted. Minimizing the number of shots required to capture all the meshes reduces the cost of photography.

Consequently, the number of photography points was reduced from 1,790 to 83 for the target building. However, time efficiency was not considered when multiple photography points were far apart, and operational issues remained because the path setting was not considered.

If we focus on the type of drone, Yoon et al. used a drone that included a VTOL model. In contrast, the others assumed a low-efficiency multirotor drone, so they were unable to pursue true efficiency in terms of photography time.

In this study, we discuss an aerial photography route for utility poles that achieves both efficiency and accuracy, using the multirotor drones or VTOL drones introduced in Chapter 2. Specifically, we determined the aerial photography course while taking supplementary shots based on the photography point that minimizes the number of shots. When the photography points are discrete, additional points are estimated, and supplementary images are captured to ensure that the accuracy of the 3D model is not compromised.

3 PROPOSAL OF THE FRAMEWORK

3.1 Proposal Summary

This study created a flight path plan based on the optimal aerial photography point that minimized the number of cameras required to achieve both accuracy and time efficiency.

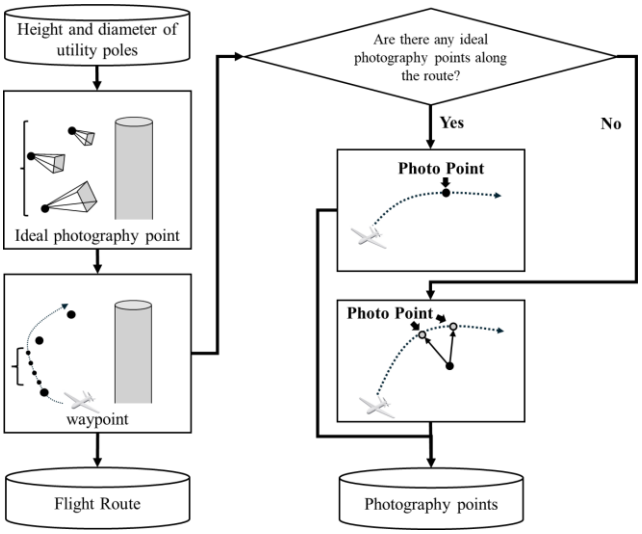


Figure 2: Proposal processing flow.

Furthermore, a path was created to perform drone aerial photography without compromising the temporal efficiency, allowing continuous flight at a constant speed. We followed this route and applied the complementary photography point-candidate extraction method described by Moritani et al. We performed interpolated photography to capture aerial shots without degrading the accuracy of the 3D model by capturing the images necessary for 3D model generation. The goal was to find the path with the shortest total aerial photography time using multiple aerial photography locations. Figure 2 shows the processing flow of the proposed method. The location data of the utility poles is used until the ideal photography point is calculated, and the information of the drone is added to calculate the waypoint. Based on the above, a flight route for generating a 3D model of a utility pole was calculated.

3.2 Flight Route Planning

The target was subdivided into smaller areas using a square mesh according to the method described by Yamazaki et al. The coordinates at which square meshes could be obtained simultaneously were determined. This can be expressed as a mathematical formula to optimize such that the binary variable x in Equation (1) is minimized. The index i represents the coordinates and orientation of a candidate photography viewpoint, while the corresponding variable x_i indicates whether this viewpoint is selected ($x_i = 1$) or not ($x_i = 0$).

$$min \sum_{\forall i \in I} x_i \quad where\ x_i \in \{0,1\} \tag{1}$$

The constraints of equation (1) are presented next. It is to extract the best shooting position for each mesh: the SfM process requires that each mesh be shot from at least two viewpoints to measure the distance by triangulation.

P_{ji} indicates whether mesh P_j (the j -th polygon) is visible from viewpoint i, with $P_{ji} = 1$ denoting visibility and $P_{ji} = 0$ indicating invisibility. The index i corresponds to the aerial

viewpoint described above, and j refers to the polygon index. This visibility determination can account not only for occlusion caused by the back side of the target object but also for occlusion by other external objects. In such cases, the visibility condition for P_{ji} must be further refined. However, this paper does not consider such extensions and leaves them as future work.

$$subject\ to \sum_{\forall i \in I} P_{ji} x_i \geq r, \forall j \in J \tag{2}$$

In each mesh, the binocular parallax of approximately 30° is represented by $A = [A_1 \cdots A_j]$. A_j is the symmetric matrix, A_{jkm} is defined as 1 if the binocular parallax between camera K and camera m with respect to the center of gravity of mesh J is approximately 30°, and 0° otherwise. Equation (3) is used as the constraint condition.

$$subject\ to \sum_{\forall i \in I} A_{jni} x_i \geq (r-1), \forall j \in J, \forall n \in I \tag{3}$$

The photography point obtained by the above equations (1)-(3) is used as a reference for the drone's flight path.

An advantage of drones is their ability to change the direction of travel without changing the direction of the aircraft through sudden movements such as turning around or forced deceleration. Therefore, constant-speed flight can be achieved by setting a waypoint with an appropriate drone turn path. Two constraint conditions were set to enable this.

The first is the upper limit of the bank angle relative to the travel direction. Specifically, the distance from the drone to the destination and the steering angle ψ_{MAX} per fixed distance L of the drone are set. Thus, the maximum bank angle φ_{MAX} from the drone point to the destination is determined.

$$\varphi_{MAX} = \frac{L}{l} \cdot \psi_{MAX} \tag{4}$$

The second constraint is the turning radius R. The turning radius R can be derived from Equation (5). V is the speed of the drone, θ is the upper limit of the bank angle derived in Equation (4), 11.26 is the gravitational acceleration converted to feet value.

$$R = \frac{V^2}{11.26 \cdot \tan\theta} \tag{5}$$

Therefore, the trajectory of the drone was revealed if it turned at the maximum bank angle. The intersection of this trajectory and the vector from the current location to the destination are determined. Assuming that the distance to an intersection is less than or equal to the distance to the destination, the destination can be reached within the maximum bank angle and is used as the waypoint. If the distance to the intersection is greater than the distance to the destination, the destination cannot be reached even at the maximum bank angle. In this case, the waypoint is the intersection of the normal vector from the current location to the destination through the destination and the locus at the maximum bank angle.

The trajectories and waypoints calculated above were used to set the route. The starting and goal points must be updated to adapt to these conditions. In this study, the starting point was set at the location of the drone. The drone movement vector was determined by connecting the coordinates of the start and drone points before rotating 180 °around the starting point. The goal point was the point next to the optimal aerial photography point that the drone had been aiming for until then.

If the vector between the drone's current location and the next optimal aerial point is $\overrightarrow{a_i}$ and the inverse vector between the drone's current location and the previous waypoint is $\overrightarrow{b_i}$, Equations (6) and (7) represent the angle formed between and the optimal aerial photography point P_i.

$$\cos\theta = \frac{\overrightarrow{a_i} \cdot \overrightarrow{b_i}}{|\overrightarrow{a_i}||\overrightarrow{b_i}|} \tag{6}$$

$$P_i = \overrightarrow{a_i} \cdot \cos(\theta - \varphi_{MAX}) \tag{7}$$

Connecting the trajectories between the waypoints calculated in these manners creates a final route plan for the drone to fly without slowing.

These settings enabled the creation of a flight path that did not slow down the drone.

3.3 Calculation of Aerial Photography Points

This section presents the aerial photography points on the calculated route. The issue with photography from the waypoint calculated in the previous section is the distance from the utility poles compared to the optimal aerial photography point of Yamazaki et al. The accuracy of a 3D model depends on image resolution. Therefore, the farther the photography point is from the object, the less accurate the 3D model becomes. To address this issue, the algorithm proposed by Moritani et al. was used to calculate the optimal additional photography point. Specifically, additional photography was performed to ensure efficient photography without compromising the 3D model quality.

Moritani et al. used two indices to find additional photography points. The first is an evaluation index for orthogonal photography. The second factor is the evaluation index of the baseline ratio. The baseline ratio is the ratio of the distance between the two images to the distance from the image to the target. These two indices are used to calculate the optimal additional photography points.

The evaluation index for positive photographs was calculated using Equation (8), where c_i represents multiple representative points between the waypoints obtained in the previous section. In this study, ten points were set between each waypoint as candidates for c_i. p is the center of gravity of each mesh when the poles are meshed, and n is the average vector of the poles at the center of gravity of each mesh. Using the above variables, we obtained an evaluation index of directly opposite photographs from each representative point in the pathway to the center of gravity of each mesh.

$$NBV^i_{frontarity} = 1 - \left(n \cdot \frac{c_i - p}{\|c_i - p\|}\right) \tag{8}$$

The baseline ratio evaluation index was calculated as follows. BH^i indicates whether the angle formed by the additional photography point candidate c_i and reference image point P_r centered at the center of gravity P of each mesh is within the specified angle. This study sets it at 45 °,

which is the same value as the standard angle of view. Although the reference image points c_r initially corresponds to the optimal aerial photography point, there are cases where it does correspond via the optimal aerial photography point. In this case, the candidate's additional photography point, located at the nearest coordinate to the optimal aerial photography point, was used as the reference image point c_r . The intermediate value between c_r and c_i is c_{ri} . According to Moritani, the recommended angle between images is 20–30 °, and 0.536 is used as a representative value for the ratio between an object and two images. Therefore, this value was used in this study.

$$BH^i = \left| \frac{\|c_i - c_r\|}{\|c_{ri} - p\|} - 0.536 \right| \quad (9)$$

$$NBV_{BH}^i = \frac{BH^i - BH_{max}}{BH_{min} - BH_{max}} \quad (10)$$

The baseline ratio obtained in Equation (9) was used in Equations (8) and (10) to calculate the evaluation index for each additional photography point candidate c_i, where w is the weight. In this study, $w_f = w_{BH} = 1/2$ to use both evaluation indices equally, according to the method described by Moritani et al. From the above, the candidate additional photography point $c_i \in C$ with the lowest value is the optimal additional photography point c_{NBV}.

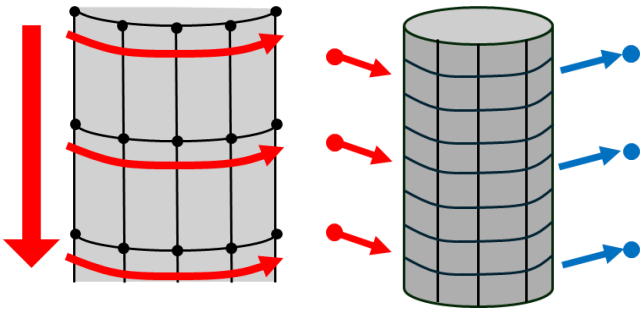
$$c_{NBV} = \underset{c_i \in C}{argmin}\left(w_f \cdot NBV_{frontarity}^i + w_{BH} \cdot NBV_{BH}^i\right) \quad (11)$$

4 EVALUATION SUMMARY

This study evaluated the compatibility between the temporal efficiency and 3D model accuracy. The accuracy of the 3D model was evaluated using the accuracy prediction index of the 3D model used by Moritani et al. Time efficiency involves the time required for aerial photography. In addition, the minimum photography point calculated using the optimal photography planning method of Yamazaki et al. was used for comparison. Yamazaki et al. assumed that the camera stops and shoots from the minimum photography point. In this evaluation, the aircraft is assumed to stop at the minimum photography point for a certain period and fly straight between the minimum photography points while accelerating and decelerating.

4.1 Parameter Settings

The poles evaluated in this study were 16-m high and 340-mm in diameter and were used as reference poles with a height of 16 m, terminal diameter of 240 mm, and former diameter of 453 mm. In this study, we evaluated the differences between the proposal and the conventional method before evaluating the attachments to the utility poles. Therefore, as a basic evaluation, we evaluated the differences between the proposal and the conventional method by evaluating the attachments and surrounding structures. In addition, the utility pole was divided into eight sections in the circumferential direction and eight in the vertical direction, totaling 64 voxels each. Therefore, the tie points were the vertices of each voxel, resulting in 72 candidate tie points.



(a)Order of tie points (b) Potential starting and finishing points

Figure 3: Parameters of utility poles

The accuracies of the tie-point candidates were evaluated. The tie points were numbered from the top of the poles, as shown in Fig. 3(a). Based on the above conditions, the degree of shape degradation at all the tie points was evaluated using the conventional and proposed methods. The speed of the drone in this evaluation varied, but the acceleration was unified at 10 km/h/s. This value was determined because the maximum acceleration of a typical aircraft is approximately 10 km/h^2. The drone's candidate starting and finishing points are at the top and bottom of the utility poles, respectively, as shown in Fig. 3(b).

4.2 Method of Evaluating Accuracy

Moritani et al.'s method was used as the evaluation value to assess the accuracy. Moritani et al. used the following six indices. This evaluation method predicts the quality of a high-density model based on the tie points that constitute the vertices of the 3D model and the camera pose of the captured image.

① Reliability F_R (Fig. 4(a))
Evaluate the number of images with visible tie points.

$$F_R(i) = |V_i| \quad (12)$$

② Average of mesh area F_a (Fig.4(b))
The average surface mesh area $A_{i,n}$ of the n th approximate object connected to the tie point i on the surface of the approximate object is denoted by $F_a(i)$ and evaluated using Equation (13).

$$F_a(i) = \frac{1}{N_{A^i}} \cdot \sum_{n=0}^{N_{A^i}} A_n^i \quad (13)$$

③ Average of mesh edges F_e (Fig.4(c))
In the approximate object surface model, the edge lengths e^i connected to the tie point i are calculated, and their average value F_e is evaluated using Equation (14), where F_e is the number of edges connected to tie point i.

$$F_e = \frac{1}{N_{e^i}} \cdot \sum_{n=0}^{N_{e^i}} e_n^i \quad (14)$$

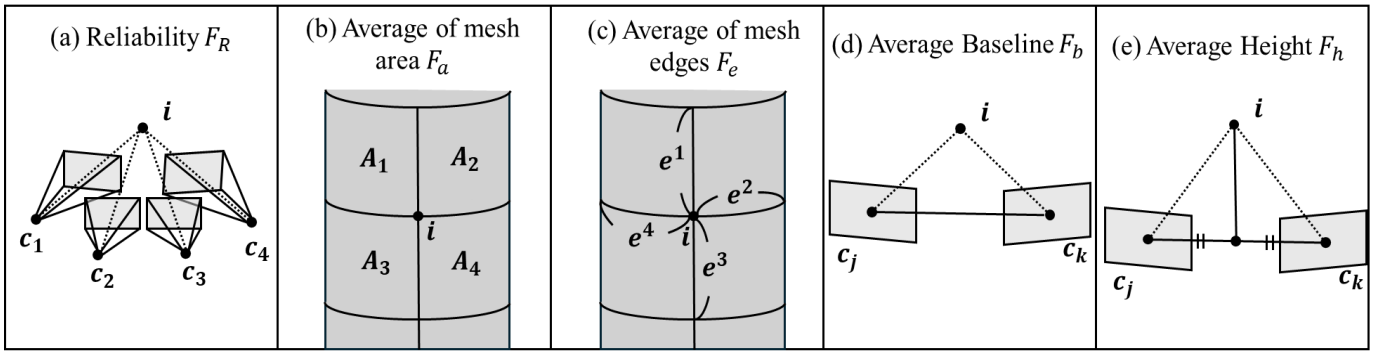


Figure 4: Summary chart of quality prediction indicators

④ Average Baseline F_b (Fig.4(d))

The distance between the centers of projection c_j and c_k of the two visible cameras j and j were calculated from the tie point i and evaluated using Equation (15), with F_b as the average value for all camera pairs.

$$F_b = \frac{1}{{}_{N_{ci}}C_2} \cdot \sum_{j=0}^{N_{ci}} \sum_{k=j+1}^{N_{ci}} \|c_j - c_k\| \quad (15)$$

⑤ Average Height F_h (Fig. 4(e))

The distance from point c_j, c_k, which bisects the baseline of two cameras j, j visible from tie point i, to tie point i was calculated and evaluated using Equation (16) with F_h as the average value for all camera pairs.

$$F_h = \frac{1}{{}_{N_{ci}}C_2} \cdot \sum_{j=0}^{N_{ci}} \sum_{k=j+1}^{N_{ci}} \|p_i - c'_{jk}\| \quad (16)$$

⑥ Base line ratio F_{bh}

The ratio of the average baseline length to the average subject distance at tie point i is F_{bh} and is evaluated using Equation (17).

$$F_{bh} = \frac{F_b(i)}{F_h(i)} \quad (17)$$

Each quality prediction index $F_X(i)(X \in R, a, e, bh)$ derived in Equations (12)-(17) is normalized using Equation (18) below and converted to $E_X(i)(\in [0,1])$.

$$E_X(i) = 1 - L \cdot (F_X - \mu_X, \sigma_X) \quad X \in [R, a, e, bh] \quad (18)$$

Here, μ_X is the mean of indicator $F_X(i)$, σ_X is the standard deviation, and $L(F_X - \mu_X, \sigma_X) = 1/(1 + \exp\left(\frac{-2(x-\mu)}{\sigma}\right))$ is the standard deviation.

$$E_{deg}(i) = \frac{E_R + E_a + E_e + E_{bh}}{4} \quad (19)$$

Finally, the average of the six energies $E_X(i)$ values was used to evaluate the degree of shape degradation, $E_{deg}(i)$, using Equation (19). $E_{deg}(i)$ has a value between 0 and 1. The higher the value, the greater the shape degradation and the less accurate the 3D model. Therefore, a value closer to 0 indicates higher accuracy of the 3D model.

4.3 Methods of Efficiency Evaluation

In this study, efficiency was evaluated based on the aerial photography time per utility pole. The procedure for calculating the specific aerial photography time varied depending on the method used. In the conventional method, the drone must stop at each photography point; therefore, acceleration and deceleration are performed between each photography point based on the acceleration of the drone. Thus, the aerial photography time between each photography point in the conventional method is $t = 2 \times \sqrt{L/a}$, where a is the acceleration.

By contrast, the proposed method requires an aircraft to fly at a constant speed of v without decelerating. Therefore, the photography time t between each photography point in the proposed method was $t = L/v$.

5 EVALUATION RESULT

5.1 Evaluation Per Utility Pole

5.1.1 Comparison Between Previous and Proposed Methods

In this section, I present the results of a comparison and evaluation of the proposed method with the conventional aerial photography method after hovering. I used a multirotor drone to perform the comparison and verification to make the two methods compatible. I compared the flight times for the shape degradation degree and speed at each tie point.

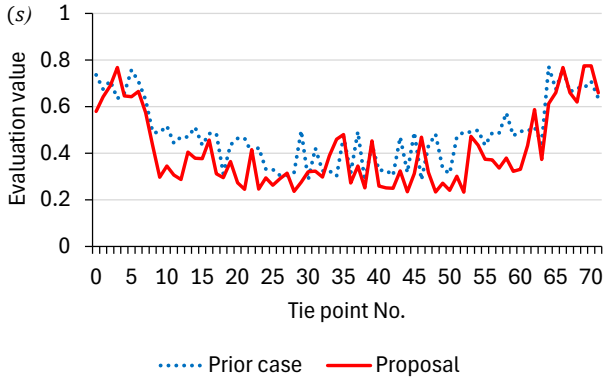
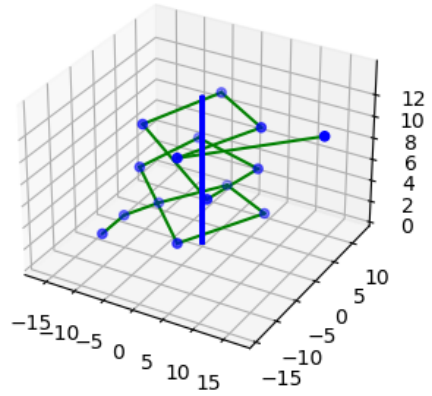
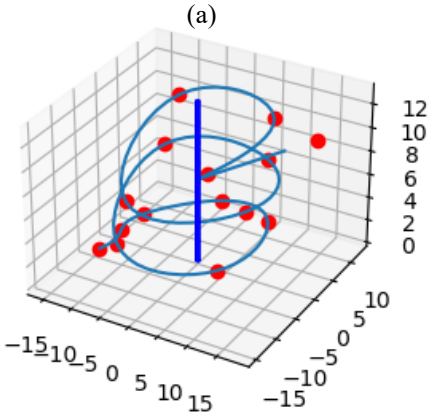


Figure 5: 3D model prediction accuracy for each tie point



(a)



(b)

Figure 6: Comparative chart of flight routes

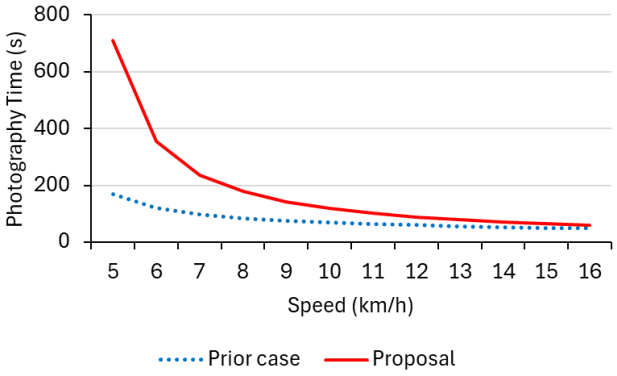


Figure 7: Relationship between speed and efficiency

As a result of the accuracy evaluation, the degree of shape degradation at each tie point is shown in Fig. 5. The average value of the shape degradation for the conventional method was 0.48, and the average value of the shape degradation for the proposed method was 0.41; thus, the shape degradation was improved by 0.07, but this was not a significant difference.

Next, the results of the efficiency evaluation are presented. The path of the proposed method using a multirotor drone is illustrated in Fig. 6(a), and the path for the conventional method is shown in Fig. 6(b). In addition, Fig. 7 shows a graph of the relationship between the upper limit of speed and aerial photography time. In the conventional method, the acceleration was set to 10 km/h; however, because the distance between the ideal photography points was not very

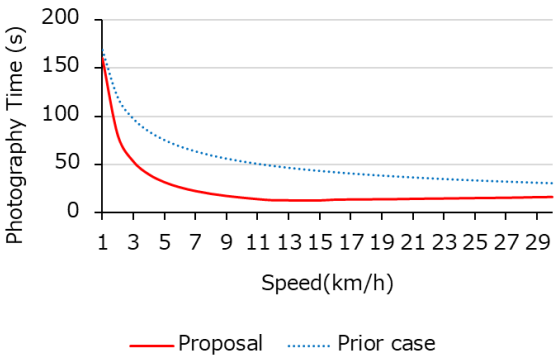
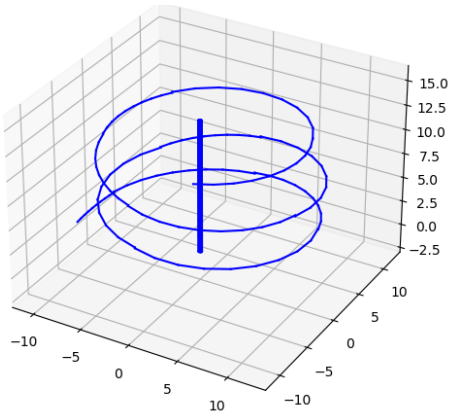
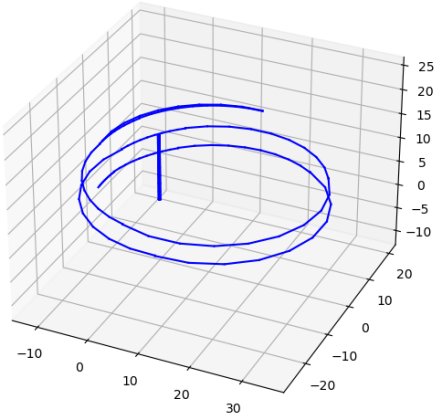


Figure 8: Speed impacts from the multirotor-type and the VTOL-type



(a)



(b)

Figure 9: 10km/h flight path and 15km/h flight path

long and was also constant, even if the upper limit of speed was increased, there was no significant impact on the aerial photography time. On the other hand, in the proposed method using a multirotor drone, the total length of the route does not change significantly even if the speed changes owing to the calculation method of the route. Therefore, as the speed increased, aerial photography time decreased. This indicates that multirotor drones can turn on a dime. In both methods, it was found that as the speed increased, the characteristics of the multirotor drones Ire reflected, and the aerial photography time converged to a similar value.

From the above, I can conclude that this algorithm is valid because there was no significant difference in either the 3D

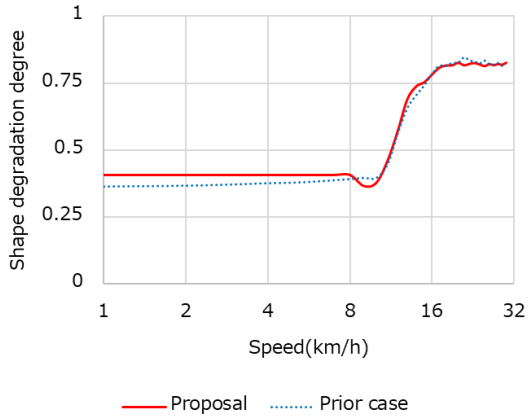


Figure 10: Relationship between speed and evaluation value

model accuracy or efficiency evaluation when the type of drone used was unified.

5.1.2 Comparison of VTOL and Multirotor Drones

In this section, we evaluate the suitability of the proposed method using VTOL-type and multirotor-type drones for aerial photography in terms of both accuracy and efficiency and describe the results of the comparison.

First, we present the results of efficiency evaluation. Figure 8 shows the relationship between the speed and total aerial photography time. The VTOL-type drone captured photographs in a shorter time than the multirotor-type drone. When the upper limit is set at 10 km/h, the result is that the VTOL type can complete filming in 27.4% of the time taken by the conventional method. However, as the upper limit of the speed increased, the difference became smaller. This was because the total distance of the aerial filming route increased as the speed increased. In addition, as a characteristic of the aerial filming route, up to a speed of approximately 12 km/h, the aerial filming route followed a circular path around the utility poles at equal intervals, as shown in Fig. 9(a). However, when it exceeds that, the aerial photography route becomes one that passes close to the utility pole on one side but traces a route that is away from the utility pole on the other side, as shown in Fig. 9(b).

Next, we examined the accuracy evaluation. Figure 10 shows the relationship between the maximum speed and average shape degradation value. The graph indicates that there was no significant variation in accuracy. This is because the aerial photography route for both the VTOL drones and multirotor drones draws a large arc as the speed increases. The farther away from the utility pole, the lower the 3D model accuracy, which is why we obtained this result.

Summarizing the above results on the temporal efficiency and 3D model accuracy, the VTOL drone showed the highest efficiency at 12 km/h. In the 3D model accuracy evaluation, the VTOL drone showed the highest accuracy at 10 km/h, although there was no significant difference from that of the multirotor drone.

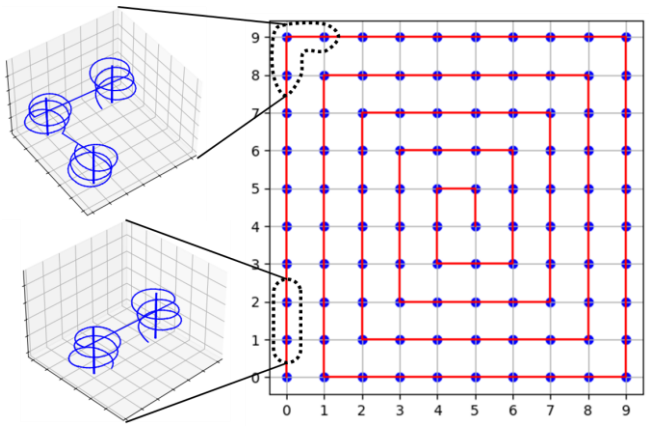


Figure 11: Manhattan model with 100 utility poles

Therefore, it was determined that using a VTOL drone at a speed of 10 km/h was the most suitable for achieving both efficiency and accuracy.

5.2 Evaluation Per 100 Utility Poles

To date, evaluations have been based on a single utility pole. We expanded the evaluation to 100 utility poles and calculated the number of drones and how long it would take to inspect all utility poles in Japan. In this verification, we calculated the pattern of VTOL drones using the proposed method, which can achieve both temporal efficiency and 3D model accuracy. The route was calculated as a Manhattan model with 100 utility poles at 40m intervals, for a total of 100, and the time for aerial photography was calculated. The speed was set at 10 km/h. In this verification, the route connecting the utility poles is set such that the angle formed by the final point of each utility pole, the route between the utility poles, and the starting point of the next utility pole are obtuse. This was added as a constraint such that the drone could fly without having to tilt the aircraft excessively. In addition, the starting point of the Manhattan model was set as the utility pole located at the corner. However, a goal was not set. This is because it is possible to set a more flexible path.

The aerial photography route in the Manhattan model is shown on Fig. 11. From the evaluation results, it can be seen that the common points of the straight and corner sections of the route are that the aerial photography route is upside-down and lined up alternately. From the results of the above verification using the Manhattan model, the total distance of the route for photographing 100 utility poles was approximately 21.7 km, and the total times for photographing them at 10 km/h were 36 min and 7 s, respectively. If one drone is operated for six hours per day, it is possible to inspect approximately 1,000 utility poles. If 100 VTOL drones are in operation, it is estimated that 35.78 million utility poles can be inspected in one year.

6 CONCLUSION

In this paper, we propose and verify a method for efficiently generating flight paths for aerial photography using drones, with the aim of reducing the labor required for inspecting utility poles in Japan. Utility poles are a critical infrastructure

supporting power and telecommunications networks; however, due to their large number and wide distribution, they are currently inspected manually. To carry out 3D model inspections of utility poles using drones, which are used for bridges and dams, it is necessary to plan the flight path and aerial photography points so that many utility poles can be photographed efficiently in a single flight.

Therefore, we devised an algorithm that aims to pass through the theoretically optimal photography position of the utility pole and searches for the optimal photography point on the main flight route while prioritizing constant-speed flight, thereby improving the model generation accuracy. Through computer evaluation, we confirmed that, compared to previous studies that assumed multirotor drones, the model accuracy of VTOL drones and this algorithm is equivalent, but flight time can be reduced by 72.6%.

Utility poles are normally accompanied by things like electrical wires, but this study only carried out a basic evaluation, so it did not take this into account. Therefore, it should be noted that this study assumes a single cylindrical object that mimics a utility pole. Future work should include evaluations with added accessories and aim to advance closer to actual field verification. On a similar note, this study did not consider the gimbal of the camera since only a basic evaluation was conducted. Since the gimbal is related to the internal factors of the camera, evaluating this can bring us closer to a demonstration experiment.

Also, although there are buildings and other structures around utility poles, this study only conducted a basic evaluation and did not take this into account. Therefore, it should be noted that the environment was ideal with no obstacles around the poles. This is one of the external factors that affect the evaluation environment. There are also other external factors related to optics, such as shadows from the utility poles themselves. Future research should focus on a more detailed subdivision of the external factors. It is necessary to evaluate each factor and approach the actual field verification. By conducting field evaluations that include these factors, it is possible to verify their practical use.

REFERENCES

[1] J. R. Gibson, “Built to Last Challenges and Opportunities for Climate-Smart Infrastructure in California”, Union of Concerned Scientists, pp.6 (2017).

[2] The WHITE HOUSE, “Fact Sheet: The President’s Plan to Make America a Magnet for Jobs by Investing in Manufacturing”, https://obamawhitehouse.archives.gov/the-press-office/2013/02/13/fact-sheet-president-s-plan-make-america-magnet-jobs-investing-manufactu (2013).

[3] U. S. Department of Transportation Federal Highway A dministration, “Highway History, In Memory of Ronal d Reagan”, https://www.fhwa.dot.gov/infrastructure/reagan.cfm (2017).

[4] Ministry of Land, Infrastructure, Transport and Tourism, “Changes in Utility Pole Phoenixes”, https://www.mlit.go.jp/road/road/traffic/chicyuka/chi_13_03.html (2024).

[5] Z. Tian, L. Jianhua, “A 3D Reconstruction Method for Pipeline Inspection based on Multi-vision”, Measurement, Vol.98, No.5, pp.35-48 (2017).

[6] H. Miyake, M. Fujiu, S. Nakayama, J. Takayama, “Evaluation of Infrastructure Conditions by 3D Model Using Drone”, 4th International Conference on Road and Rail Infrastructure, Vol.9, No.3, pp.597-602 (2016).

[7] Ministry of Land, Infrastructure, Transport and Tourism, “Securing Infrastructure Development Leaders, Improving On-Site Productivity, Introducing New Technologies, etc.”, https://www.mlit.go.jp/hakusyo/mlit/h27/hakusho/h28/html/n1321000.html(2024).

[8] T. Ishida, K. Nakata, J. Okui, S. Yamada, “Verification of a method for surveying river basins using VTOL drones in emergency situations (in Japanese)”, Proceedings of the 2024 Japan Society of Erosion Control Research Meeting, No.19, pp. 323-324 (2024).

[9] D. Kubo, “History and Technological Development of Unmanned Aircraft Systems (Drones) (in Japanese)”, Measurement and Control: journal of the Society of Instrument and Control Engineers, Vol.56, No.1, p.12-17(2017).

[10] J. B. Sankey, S. Ravi, C. S. A. Wallace, R. H. Webb, T. E. Huxman, “Quantifying Soil Surface Change in Degraded Drylands: Shrub Encroachment and Effects of Fire and Vegetation Removal in a Desert Grassland”, Journal of Geophysical Research: Bio geosciences: Vol.117, Issue G2(2012).

[11] D. Moon, S. Chung, S. Kwon, J. Seo, J. Shin, “Comparison and Utilization of Point Cloud Generated from Photogrammetry and Laser Scanning: 3D world model for smart heavy equipment planning”, Automation in Construction, Vol.98, Pages 322-331(2019).

[12] S. Yoon, G. Gwon, J. Lee, H. Jung, “Three-Dimensional Image Coordinate-Based Missing ROI Area Detection and Damage Localization for Bridge Visual Inspection Using Unmanned Aerial Vehicles”, Structural Health Monitoring, Vol.20, No.4, pp.1462-1475 (2020).

[13] R. Moritani, S. Kanai, H. Date, Y. Niina, R. Honma, “View Planning for High-quality and Efficient as-is Model Reconstruction by SfM-MVS (6th report) ― Estimation of the Additional Camera Poses Based on The Extracted Low-Quality Regions ―(in Japanese)”, 2020 Journal of the Japan Society of Precision Engineering, No.22, pp.43-44 (2020).

[14] K. Yamazaki, K. Okahara, A. Minesawa, “Proposal of an Optimal Imaging Planning Method for 3D Reconstruction Using Geospatial Information (in Japanese)”, IPSJ Symposium series DICOMO Multimedia, Distributed, Cooperative, and Mobile Symposium, (DICOMO 2023), pp.611- 616 (2023).

[15] H. Dan, N. Inazu, T. Ozaki, S. Kubota, and Y. Yasumuro, “Optiminal ff UAV Path Planning for Aerial Photography Survey of Outdoor Structures”, Japanese Journal of JSCE F3, Vol.74, No.2, pp.159-166 (2020).

Koji Yamagishi graduated from the Faculty of Information Frontier, Kanazawa Institute of Technology. He then entered the Graduate School of Engineering at Kanazawa Institute of Technology. He is a member of the Information Processing Society of Japan.

Yuichi Tokunaga received a Ph.D. degree in science and engineering from Ritsumeikan University in 2009. He joined Mitsubishi Electric Corporation in 1990 and engaged in R&D of the high-reliability computer, wireless sensor network, the algorithms of time-synchronization and pointing, the network protocol for industrial applications, and data analytics for CBM (Condition Based Maintenance). He has been a professor at Kanazawa Institute of Technology since 2019. He is a member of the ITS steering committee of IPSJ and a member of IEEE and ISCIE.

Regular Paper

Efficient Classification of Non-Functional Requirements Using ChatGPT's Function Calling Feature

Kazuhiro Mukaida*, Seiji Fukui** , Takeshi Nagaoka** , Takayuki Kitagawa ** , Shinpei Ogata* and Kozo Okano*

*Graduate School of Science and Technology, Shinshu University, Japan
**Toshiba Corporation, Japan
24hs254c@shinshu-u.ac.jp, {Fukui.Seiji, Nagaoka.Takeshi, Kitagawa.Takayuki}@toshiba-sol.co.jp, {ogata, okano}@cs.shinshu-u.ac.jp

Abstract - Efficient analysis of requirement specifications is crucial for improving their quality, which is important in software development. We focus on non-functional requirements (NFRs), which are often overlooked in requirement definitions of system developments and propose a method that allows individuals without extensive expertise to efficiently extract and classify NFRs from requirement specifications. The proposed method aims to efficiently enhance the quality of software requirement specifications by enabling the extraction and classification of NFRs with minimal expertise.

Previously, the authors experimented with creating dedicated deep learning models for classification tasks and then used pre-trained Transformer models like BERT and GPT-2, trained on human-annotated datasets. However, recent advancements, such as tools like ChatGPT, enable classification via prompt interactions alone. In this paper, we explore the capabilities of ChatGPT's Function Calling feature, utilizing an approach that optimizes its behavior, aiming to demonstrate its superior classification performance compared to prompt-only responses and conventional classification methods, which require large training datasets.

Function Calling significantly reduced ambiguities and improved classification accuracy by ensuring adherence to predefined classification boundaries. For example, GPT-4 Turbo demonstrated an F1 score improvement from 0.681 to 0.753, and GPT-3.5 Turbo achieved an increase from 0.587 to 0.651. Additionally, GPT-4o showed a gain from 0.754 to 0.780. These improvements highlight the practical utility of Function Calling as a primary classification tool.

As supplementary verification, we conducted two additional analyses. Fine-tuning GPT-3.5 Turbo on small datasets significantly enhanced its performance, achieving an F1 score of 0.796. Similarly, incorporating sentence concatenation by linking preceding and following sentences improved contextual understanding, increasing accuracy from 0.792 to 0.831. These approaches, while complementary, further validated the robustness of Function Calling for NFR classification tasks.

Future research should address remaining challenges, such as improving the model's contextual understanding and developing targeted training datasets that emphasize the most challenging classification categories. These findings highlight the potential of advanced natural language models like ChatGPT in making NFR classification more efficient and precise.

Keywords: Function Calling, ChatGPT, GPT-4o, Non-Functional Requirements, Documents Classification

1 INTRODUCTION

In information system development, requirements are broadly divided into functional and non-functional requirements [1]. Functional requirements define specific functions that the system must perform, whereas non-functional requirements describe the overall qualities of the system, such as availability, performance, reliability, and efficiency. These non-functional requirements define expectations for the system's actual operating environment and are crucial for ensuring the overall quality of the software [2].

1.1 Importance and Challenges of Non-Functional Requirements

Non-functional requirements are critically important for the success of a system, but their abstract nature makes their identification and classification challenging. If non-functional requirements are not adequately defined, they can severely impact system performance, usability, and security. Therefore, clearly defining non-functional requirements and managing them throughout the development process is one of the major challenges in system development. Insufficient non-functional requirements can result in the system failing to meet expected performance, ultimately leading to decreased user satisfaction.

Non-functional requirements are essential for ensuring the overall quality of the system. Characteristics such as system response time, throughput, availability, security, usability, and scalability directly affect user experience and reliability. For example, if a system has a slow response time, users may find it difficult to use and avoid using it altogether. Additionally, if security requirements are not adequately met, the system may be vulnerable to external attacks, increasing the risk of important data being leaked. Thus, non-functional requirements are vital for maintaining the health of the system and the trust of users.

To properly define and agree on non-functional requirements, it is necessary to clearly identify and classify

them. However, since non-functional requirements are abstract and diverse, manual classification is time-consuming, labor-intensive, and requires specialized knowledge. Disparities in understanding non-functional requirements between users and vendors often arise, becoming an obstacle to the success of system development.

1.2 Emergence and Advancement of Automated Classification Techniques

As a means of addressing the challenges of identifying and classifying non-functional requirements, automated classification techniques are gaining attention. In particular, research on document classification techniques using deep learning and large language models (LLMs) has been actively conducted. Since the advent of Transformer by Vaswani et al. [3], models such as BERT (Bidirectional Encoder Representations from Transformers) by Devlin et al. [4] and GPT (Generative Pre-trained Transformer) by Radford et al. [5], which emerged in 2018, have been pre-trained on large amounts of text and fine-tuned for specific tasks to improve classification accuracy. The authors have also experimented with automatic classification using a model incorporating BERT and GPT-2 [6] in the previous paper [7].

Given the abstract nature and diversity of non-functional requirements, the introduction of automated classification techniques is highly beneficial. First, defining and classifying non-functional requirements require the expertise and time of experienced professionals, making manual classification costly and time-consuming. Automated classification techniques significantly reduce these manual efforts, allowing for efficient classification. Moreover, automated classification provides consistent results and reduces human errors and biases. Additionally, accurate and swift classification of non-functional requirements enables appropriate requirements management from the early stages of a development project, thereby improving the project's success rate. For instance, properly classifying and recognizing system performance requirements early on can help prevent performance issues in later development stages.

Recently, with the widespread use of conversational models like ChatGPT, inference through prompt interactions is becoming possible. Conversational LLM models generate appropriate responses in response to questions and instructions during interaction with users in natural language. This conversational capability allows users to perform advanced inferences and information searches without requiring specific knowledge or skills.

Models generate responses based on prompts. Thus, by designing prompts appropriately, the output of the model can be controlled to obtain responses suitable for the intended purpose. For example, the format and content of prompts can be adjusted to handle various tasks, such as question answering, text generation, summarization, and translation.

However, there are limitations to performance in classification tasks. Classification tasks involve categorizing text or data into specific categories or labels, which is one of the fundamental applications of deep learning models. For instance, detecting spam emails or categorizing product reviews as positive or negative are common classification tasks. While the classification performance of conversational models has improved, they may still lag behind dedicated traditional models for certain complex tasks or large datasets. Particularly, when handling multi-class classification beyond binary classification, hallucinations may occur, resulting in the unintended creation of classification categories, which negatively impacts accuracy.

In this paper, we attempt to classify non-functional requirements using the Function Calling feature of ChatGPT. Function Calling is a means of obtaining additional information by calling other APIs or functions during prompt interactions, allowing integration with external services and databases. Generally, by using this feature, more advanced processing and data retrieval become possible.

This paper focuses on utilizing arguments required for calls during the Function Calling process as classification data. By employing enum to restrict and enumerate variable types, unintended category creation due to hallucinations can be eliminated, ensuring that the classification remains within the intended boundaries. The model automatically suggests appropriate arguments based on contextual information, and proper argument content is directly linked to accurate calls and responses. Utilizing these argument suggestions as classification data is expected to improve the classification accuracy of non-functional requirements.

In addition to the core focus on the outcomes of Function Calling, the following two experiments were supplementary investigations aimed at further enhancing the effectiveness of the Function Calling feature. While these experiments provide valuable insights into potential improvements, the primary objective of this paper remains centered on the results derived from Function Calling.

First, we verify the impact of small-scale fine-tuning using non-functional Requirements Grades [8] on classification by Function Calling. By optimizing the Function Calling feature of ChatGPT based on clearly defined standards using non-functional Requirements Grades, we aim to improve the classification accuracy of non-functional requirements. The outcomes of this study are expected to contribute to the development of automated classification technologies for non-functional requirements and enhance the efficiency of the quality assurance process in system development.

Second, we propose a classification method that uses sentence concatenation to provide additional contextual information. In conventional classification methods, only the target sentence is classified, but incorporating contextual information can improve classification accuracy. In requirement specifications, similar non-functional requirements may span multiple sentences, and the preceding and following sentences are often useful as additional contextual information for the target sentence. In this method, we embed the test data, determine similarity based on cosine similarity between the target sentence and its surrounding sentences, and concatenate related sentences before classification. This allows for classification that

considers contextual information, thereby improving classification accuracy.

In the future, automated classification technologies for non-functional requirements are expected to evolve further, enabling more accurate and efficient classification. In particular, with the advancement of large language models, it will become possible to understand the abstract nature of non-functional requirements more deeply and perform classification accordingly.

2 RELATED RESEARCH

2.1 Classification of Non-Functional Requirements

Kinoshita et al. [9] proposed a method for extracting non-functional requirements (NFRs) from Japanese requirement documents. Specifically, they established keywords related to NFRs and applied them to the documents to effectively identify relevant requirements. Additionally, they introduced an approach for detecting errors in NFRs by defining case frames for the actions described in requirement sentences. Their method involves validating these sentences using criteria such as omission, ambiguity, redundancy, and inconsistency, thereby enhancing the overall quality of the requirements.

2.2 Classification Using Deep Learning

Gnanasekaran et al. [10] researched and developed recurrent neural network (RNN) models, known for their effectiveness in processing sequential natural language text. They conducted experiments using these models to classify NFRs described in natural language into five categories: maintainability, operability, performance, security, and usability. The experimental validation was based on two datasets encompassing approximately 1,000 NFRs, demonstrating the potential of RNNs in accurately classifying NFRs.

Kitagawa and Nagaoka [11] proposed an automatic classification method using Word2Vec and convolutional neural networks (CNN). Their approach extended the classification targets beyond NFRs and project management-related sentences to include functional requirements and other types of sentences. They aimed to automatically classify all sentences in Japanese Request for Proposal (RFP) documents, improving the efficiency and accuracy of requirement specification analysis.

2.3 Classification Using Large Language Models

Since the introduction of the Transformer architecture, significant advancements have been made in natural language processing tasks using LLMs based on Transformers. These models have largely replaced traditional recurrent neural networks such as RNNs, LSTMs, and GRUs. By fine-tuning large pre-trained models like BERT [4] for specific tasks, researchers have achieved models optimized for those tasks, often attaining higher accuracy.

Zhu et al. [12], starting from BERT, introduced a novel method to enhance classification performance by formalizing input sentences as natural language templates and leveraging knowledge expansion [13]. This approach improved short text classification by integrating additional contextual information into the prompts.

Gutierrez et al. [14] reported that fine-tuned BERT models outperformed interactive GPT-3 models in tasks such as named entity recognition and relation extraction within the medical domain. Similarly, Sun et al. [15] found that while interactive models like GPT-3 can perform text classification tasks, fine-tuned task-specific models still achieve higher accuracy due to their specialized training.

Ibe et al. [16] conducted experiments using large generative language models to automatically classify requirement specifications with few examples by adjusting prompts. Their results indicated that BERT-based classifications outperformed interactive models like ChatGPT in terms of accuracy, highlighting the effectiveness of task-specific fine-tuning.

On the other hand, the performance of ChatGPT as a question-answering system (QAS) has been reported to match or even exceed that of traditional task-specific QAS models. Tan et al. [17] demonstrated that ChatGPT provides superior robustness and explainability in conversations compared to traditional QAS, offering enhanced user interaction and adaptability.

Brown et al. [18] evaluated GPT-3's performance in few-shot learning settings, reporting high effectiveness in natural language processing tasks under zero-shot and one-shot conditions. Their work highlighted the capability of large language models to perform various tasks without extensive task-specific fine-tuning, showcasing the potential of models like GPT-3 in few-shot learning scenarios.

In studies focusing on prompt engineering, the importance of In-Context Learning (ICL) has been emphasized. Dong et al. [19] provided a comprehensive overview of ICL, demonstrating its applicability across diverse tasks and its ability to improve model performance by incorporating contextual information during inference. Liu and Yang [13] further explored knowledge-enhanced prompt learning for few-shot text classification, showing that integrating external knowledge can significantly boost performance.

Min et al. [20] analyzed how the selection of demonstrations in ICL affects model performance, clarifying the optimal conditions for demonstration selection to maximize effectiveness. Wan et al. [21] showcased the effectiveness of ICL in relation to extraction tasks, proposing methods with higher flexibility and accuracy compared to conventional approaches by utilizing contextual cues within the prompts. These studies highlight the flexibility and performance of interactive models and ICL across various tasks, opening new possibilities in the field of natural language processing.

3 PROPOSED METHOD

This paper proposes an automatic classification method for NFRs in system development specification documents. This method leverages the Function Calling feature of GPT models, a type of large language model, and is based on IPA standards.

3.1 Function Calling

Function Calling is a feature provided by the API of large language models, enabling the model to indirectly interact with external APIs or systems and generate specific actions. This technology involves parsing natural language queries, selecting appropriate functions, and generating JSON responses. For example, in response to the query “Tell me about the weather in Tokyo,” the model suggests calling a function that uses a weather API with the region name as an argument. The argument "Tokyo" is prepared for the weather API call. The program then retrieves Tokyo's weather through the function and weather API, then passes this information back to the model as an additional prompt. The model then provides the final answer, significantly expanding the potential for interactive applications and services using the model. This feature bridges the gap between natural language understanding and the execution of specific actions, enhancing the model's practicality and allowing it to flexibly respond to user requests. Function Calling not only streamlines the interaction between the model and external systems but also improves the overall efficiency of handling user queries by automating the process of converting natural language into actionable commands.

3.2 Classification using Function Call

A derivative use of Function Calling involves extracting structured data from text. For example, a function like extract_data(name: string, birthday: string) can be defined and invoked as needed, extracting a person's name and birthday from the text as arguments. This functionality allows for the rapid and accurate extraction of necessary information from large volumes of text data, facilitating efficient data analysis and information management. This method provides a more structured and organized approach to handling text data, ensuring that the extracted information is consistent and accurate.

Function Calling can also help in selecting the most appropriate argument based on the overall context of a query, even if the specific term is not explicitly mentioned. By leveraging the model's natural language understanding, it can infer the intended meaning and relevant information from the user's input.

In this study, rather than using Function Calling for extracting structured data from documents, we applied it to document classification. Normally, Function Calling in LLMs is triggered only when the model determines it is necessary. However, we configured parameters to ensure that Function Calling is always invoked whenever a classification target sentence is input. This guarantees that Function Calling is consistently performed during classification tasks.

The function invoked in this process exists only as a description of its name and outline and does not exist as an actual implemented program. The primary purpose of Function Calling in our method is to return arguments that indicate classification categories. These arguments are subsequently used within the program for further processing. Since the goal of this experiment is to acquire the necessary arguments for classification, the Function Calling process terminates once the arguments are obtained, without invoking any external modules or APIs. This ensures that the classification process benefits from the structure and consistency of Function Calling without relying on external integrations.

To achieve classification, we enforced the invocation of a dummy function through prompts containing evaluation sentences. The model suggests classification categories as arguments during this process. By specifying the argument type as enum, we enumerated the classification categories, limiting suggestions to predefined options and preventing hallucinations. This approach automates and streamlines the classification process, reducing the potential for human error and enhancing overall efficiency.

3.3 Supplemental Analysis 1 Fine Tuning for GPT-3.5

Section 3.3 and 3.4 aim to enhance the results achieved by utilizing Function Calling through two additional techniques. The purpose of these experiments is to improve the classification accuracy of NFRs by fine-tuning and enhancing contextual understanding.

Firstly, this paper attempts to improve models with classification accuracy by additional training of GPT-3.5 Turbo with a small amount of training data. Fine-tuning enables the model to learn language expressions specific to NFRs tasks or domains. This helps the model understand specialized terminology and contexts that are challenging for general language models, achieving higher classification accuracy. Moreover, fine-tuning with a small amount of training data significantly reduces the time and cost associated with preparing annotated large datasets, compared to traditional methods. For complex tasks like NFRs classification, it is crucial to learn efficiently from limited examples.

Fine-tuning not only tailors the model to specific tasks but also enhances its ability to handle nuanced and domain-specific language, resulting in more precise and reliable classifications. This approach is feasible even in scenarios where annotated data is scarce, enabling effective learning and adaptation with minimal resources.

3.4 Supplemental Analysis 2 Enhancing Contextual Understanding

Lastly, this paper proposes the use of sentence concatenation to provide additional contextual information that improves classification accuracy. The classification model has traditionally focused only on the target sentence,

but incorporating contextual information can enhance performance by capturing the broader context in which the target sentence appears. In requirement specifications, sequential sentences often describe similar non-functional requirements, and the preceding and following sentences frequently contain valuable information that helps clarify the classification of the target sentence.

In this method, embeddings are generated for each sentence in the test data using the text-embedding-3-large model. This model, provided by OpenAI, is an advanced natural language processing model capable of converting sentence semantics into numerical vectors. The embeddings are 3072-dimensional vectors that capture the nuances of sentence content and context, enabling accurate similarity assessments. These representations are essential for various tasks, including classification, search, and semantic comparison.

To identify relevant context, cosine similarity is calculated between the target sentence and its preceding and following sentences. Cosine similarity measures the angle between two vectors, with values closer to 1 indicating higher similarity. When the cosine similarity exceeds 0.5, the adjacent sentence is considered contextually relevant and is concatenated with the target sentence. This process generates a new, enriched input that reflects not only the target sentence but also the surrounding context. For example, if there are consecutive sentences A, B, and C, and B is the classification target, and if A and B are similar while B and C are not, the classification is performed using a concatenated sentence consisting of A and B.

It is possible that increasing the amount of text through concatenation may enhance classification performance to some extent, even if the added text is not directly relevant. However, adding unrelated text can sometimes introduce noise and ambiguity, which may complicate classification rather than improve it. To address this, sentences are concatenated selectively based on their contextual relevance, as determined by cosine similarity. This approach helps ensure that the additional information enhances the classification process by reinforcing the semantic continuity of the text.

Table 1 Sources of Training Data by Organization

Organization	Project Description	Date
Ministry of Economy, Trade and Industry	Industrial Safety System Update	Jan-22
Digital Agency	Development and Operation/Maintenance Work for Account Information Registration Linkage System	Jul-22
Ministry of Health, Labor and Welfare	National Unified System for Medical Function Information Provision System and Pharmacy Function Information Provision System,	Jan-23

By concatenating relevant sentences, the model gains supplementary cues that contribute to more accurate classification. This method helps reduce the risk of misclassification by avoiding the inclusion of unrelated information and maintaining a focus on contextually aligned data.

The test data described in Section 4.1, derived from actual requirement specifications, was used for this experiment. Since the target sentences were extracted directly from these documents, the preceding and following sentences were naturally part of the same text, ensuring contextual relevance.

Experiments conducted with and without sentence concatenation demonstrated that improvements in classification accuracy were due to the inclusion of relevant contextual information, rather than merely increasing the amount of text. This structured approach highlights the value of selectively expanding context to enhance classification performance while minimizing the risk of unnecessary complexity.

4 EVALUATION EXPERIMENT

4.1 Models and Test Data Set

In this paper, we utilized the ChatGPT API with the following models:

GPT-4 Variants
- gpt-4o-2024-05-13
- gpt-4-turbo-2024-04-09

GPT-3.5 Turbo
- gpt-3.5-turbo-0125

The gpt-4o-2024-05-13 model represents the advanced current iteration of the GPT-4 series, incorporating the recent advancements in language understanding and generation capabilities.

The gpt-4-turbo-2024-04-09 model is offering optimized performance and a popular choice for a wide range of practical applications.

The gpt-3.5-turbo-0125 model, although from a previous generation, provides a valuable comparison point. It offers cost-effective performance and promptly supports fine-tuning, allowing for customization to specific tasks or domains. This makes the gpt-3.5-turbo-0125 model advantageous for scenarios where budget constraints and the need for tailored solutions are paramount.

Table 2 Number of Test Data Instances for Each Classification

Availability	81
Performance/Scalability	70
Operability/Maintainability	90
Migratability	60
Security	90
System Environment/Ecology	23
Total	414

In our analysis, we specified that the argument type for function calling properties is enum. For enum types, we defined "availability, performance/scalability, operability/maintainability, migratability, security, and system environment" based on the classification of non-functional requirements grades. By doing so, we were able to prevent hallucinations of unspecified categories.

Models provided responses for each evaluation sentence by sentence. The temperature was set to 0 to ensure deterministic outputs.

The primary focus of the experiment is to verify the superiority of using Function Calling capabilities over the baseline case, which relies solely on prompt interactions for classification. At the same time, we are able to observe the differences between models. This comparison provides valuable insights into their practical applications and efficiency, highlighting the strengths and possible improvements.

The test dataset used in this paper was constructed based on the existing requirement specifications of public tenders announced by government agencies shown in Table1. The requirement specifications were randomly selected. The test data was randomly extracted from these documents, and labeling was performed by us based on the IPA's non-functional requirements grades. The detailed distribution of sentences for each label is shown in Table 2.

4.2 Baseline Classification Method without Function Calling

To establish a baseline for classification accuracy, we evaluated a method that relies solely on prompt interactions, without utilizing the Function Calling feature. This baseline approach involved presenting the model with a prompt that defined the classification task and the sentence to be classified. The experiment was conducted in a zero-shot setting, with no examples provided to the model. This approach was chosen to evaluate the model's performance without prior contextual learning or fine tuning, thereby establishing a fair baseline for comparison.

The prompt used for this baseline is as follows:

"In the IPA's non-functional requirement grades, non-functional requirements are classified into six categories: availability, performance/scalability, operability/maintainability, migratability, security, and system environment. Answer with the name of only one of these six categories."

This prompt was consistently applied across both the prompt-only method and the Function Calling method to ensure a fair comparison. The purpose of the experiment was not to argue that the Function Calling method inherently surpasses the performance of the prompt-only method under varying conditions. Instead, the objective was to demonstrate that the addition of Function Calling to the same prompt leads to improved classification accuracy. By applying the same prompt in both cases, the comparison highlights the effect of integrating Function Calling, rather than differences in prompt design.

It is acknowledged that the application of advanced prompt engineering techniques or in-context learning could potentially enhance the performance of the prompt-only method. However, such approaches fall outside the scope of this study. The primary focus is to illustrate how the incorporation of Function Calling enhances performance, even when used alongside a standard, unoptimized prompt.

By establishing this baseline, the experiment aims to provide a clear and reproducible comparison between the two approaches, ensuring that the observed improvements in classification accuracy can be attributed directly to the inclusion of Function Calling, rather than variations in prompt formulation or tuning strategies.

4.3 Effect of Function Calling

Table 3 compares the classification results using Function Calling and prompt-only methods for each model. The evaluation metrics include accuracy, precision, recall, and F1-score, with the averages calculated using macro-averaging.

In most models, the method using Function Calling achieved higher accuracy compared to the prompt-only method. For instance, GPT-3.5 Turbo shows an accuracy of 0.705 with Function Calling versus 0.696 with prompt only. GPT-4 Turbo demonstrates even more significant gains with Function Calling, achieving 0.775 compared to 0.740 with prompt-only. The GPT-4o model also exhibits superior performance with Function Calling, reaching an accuracy of 0.792 compared to 0.783 with prompt only. For GPT-4o, the performance improvement with prompt-only methods is remarkable, reducing the advantage of Function Calling.

Table 3 Classification Performance Metrics for GPT-3.5 Turbo, GPT-4 Turbo, and GPT-4o using Function Calling and Prompt-Only Methods

	GPT-3.5 Turbo		GPT-4 Turbo		GPT-4o	
	Function Calling	Prompt only	Function Calling	Prompt only	Function Calling	Prompt only
Accuracy	0.705	0.696	0.775	0.740	0.792	0.783
Precision	0.719	0.608	0.811	0.766	0.785	0.787
Recall	0.648	0.618	0.734	0.674	0.779	0.741
F1-score	0.651	0.587	0.753	0.681	0.780	0.754

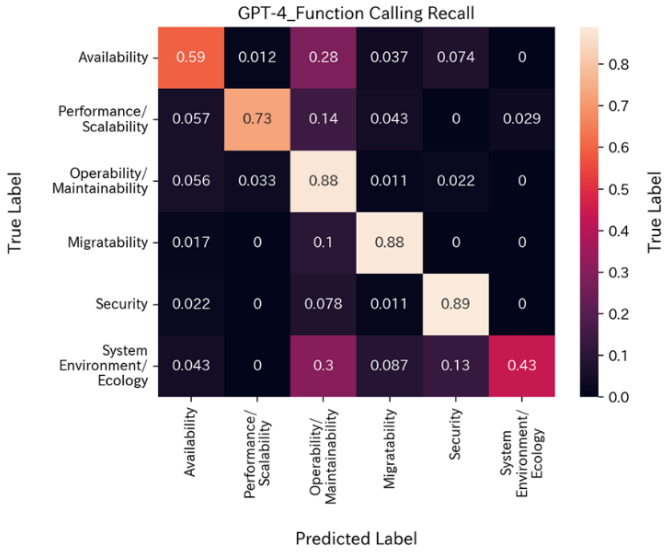


Figure 1 Recall Matrix for GPT-4 Turbo Function Calling and Prompting Performance Across 6 Categories

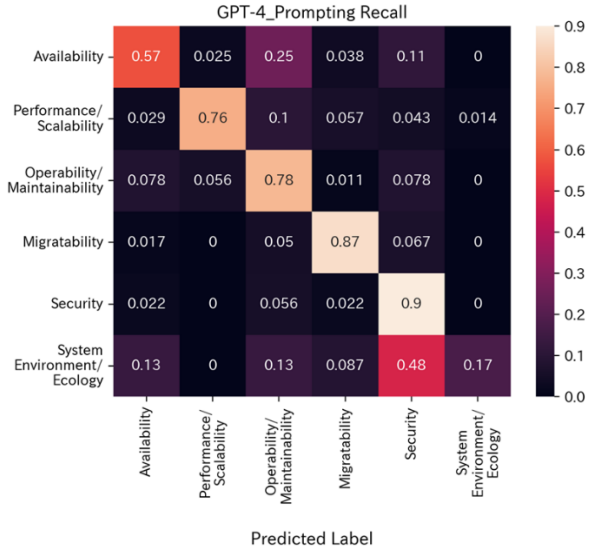


Figure 2 Recall Matrix for GPT-4 Turbo Prompting only Performance Across 6 Categories

Overall, these results suggest that the use of Function Calling significantly enhances the performance of language models across various evaluation metrics. While there are some nuances, particularly with GPT-4o, the general trend highlights the practical advantages of this approach in classification tasks.

4.4 Misclassification Analysis

In this section, we examine the misclassification issues observed across various categories, focusing on the "System Environment/Ecology" category. Figure 1 and Fig. 2 illustrate the recall values for Function Calling and prompting methods across categories, highlighting areas with pronounced misclassification.

Both Function Calling and prompting methods demonstrate low recall values for the "System Environment/Ecology" category, with prompting performing particularly poorly, likely due to the inherent complexity of this category. To illustrate the challenges and successes in classifying non-functional requirements, we present specific examples of correctly classified and misclassified cases within the "System Environment/Ecology" category.

The first example pertains to the System Environment category, "The construction environment for the next-generation security network assumes the use of the cloud, making physical servers and devices unnecessary." It was correctly classified as belonging to the System Environment category due to its explicit focus on operational premises and infrastructure design.

Conversely, the second example highlights a misclassification case. The requirement "Access to the facility and rooms housing managed devices is restricted, particularly for rooms containing devices, where only authorized personnel with management privileges are permitted entry" was incorrectly categorized as Security instead of System Environment. While this requirement involves access control measures, its primary focus is on the physical environment and the placement of managed devices, which aligns more closely with the System Environment category.

According to the IPA's non-functional requirements grade, the System Environment/Ecology category encompasses system constraints/premises established at installation, system characteristics, compliance standards, and environmental . associated with users and the region surrounding the system. Collectively, these elements represent the broader environmental and contextual factors influencing system operation and integration.

The low recall in this category likely stems from its broad and multifaceted nature, covering diverse aspects such as technical constraints and user characteristics. This complexity introduces ambiguity, making it difficult for models to distinguish relevant terms from other categories. Additionally, overlap exists between the "System Environment/Ecology" category and others; for instance, environmental conditions may also pertain to security concerns. Such overlaps complicate classification, as models often struggle to differentiate between closely related concepts. Moreover, interpreting terms related to the system environment is highly context-dependent, which hinders the models' ability to generalize accurately.

Overall, misclassification remains a critical challenge, particularly within the "System Environment/Ecology" category, due to its inherent complexity and the overlap with other classifications.

Table 4 Number of Training Data Instances for NFRs Grade

Availability	168
Performance/Scalability	175
Operability/Maintainability	158
Migratability	101
Security	163
System Environment/Ecology	168
Total	933

5 SUPPLEMENTAL EXPERIMENTS

5.1 Training Data for Fine Tuning and Embedding Analysis Fine Tuning Models

GPT-3.5 Turbo is capable of fine-tuning with even small amounts of data. Fine-tuning is greatly influenced by the quality of training dataset used, making its selection extremely important. In this paper, we utilized descriptions of non-functional requirement grades as the training dataset instead of using sentences from actual specifications. The primary reason for this choice is that it allows for more efficient annotation, as the classification of sentences is more straightforward. Additionally, one reason is the hypothesis that it may be difficult to learn the diversity of NFRs with a small amount of data with small numbers of training data. The non-functional requirement grades classify NFRs into six categories: availability, performance/scalability, operability/maintainability, migratability, security, and system environment/ecology. Descriptions related to each category were labeled with the respective classification and used as training data. As shown in Table 4, the training dataset comprises a total of 933 instances, distributed across the six NFR categories.

For each experiment, a necessary number of instances was randomly extracted from this dataset. This approach ensured that the model could be fine-tuned effectively while utilizing a representative sample of the data for each specific experiment. Figure 3 shows the t-SNE plot of descriptions for each non-functional requirement grade after embedding them using the text-embedding-3-large model provided by OpenAI, which converts the text into 3072-dimensional vectors. Each color represents one of the six NFR categories. The clear clustering of some categories, like security, suggests that the classification task for these categories might be relatively straightforward for the model, likely resulting in higher accuracy, precision, and recall. Conversely, the dispersion observed in some categories, like the system environment/ecology categories, indicates potential challenges by enhancing context understanding. in achieving high classification performance. These categories might require more sophisticated models or additional context to improve classification accuracy. The t-SNE visualization serves as a validation tool, demonstrating that the embedding model can capture and represent the similarities and differences among the NFR descriptions to a

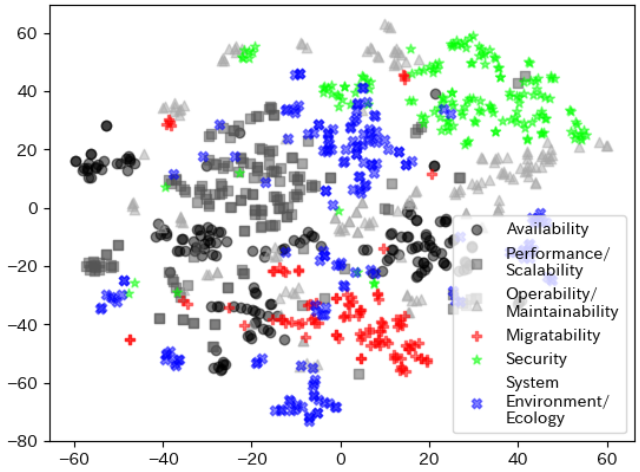


Figure 3 t-SNE Visualization of Embedded Training Data

significant extent. This visual validation supports the choice of using for fine-tuning and classification tasks.

In conclusion, the t-SNE plot effectively illustrates the clustering behavior of the NFR descriptions based on their embeddings. The distinct clusters for some categories and the overlapping regions for others provide valuable insights into the complexity of the classification task. These insights highlight the strengths and potential context understanding in using the current model for classifying non-functional requirements, guiding future improvements, and fine-tuning efforts.

5.2 Fine Tuning Effect

In order to comprehensively evaluate the progression of loss and its impact on accuracy, three different fine-tuning models were tested, each designed to address varying conditions in terms of dataset size and number of epochs, thus providing a robust analysis.

In finetuning 1, illustrated in Fig. 4-1, training was conducted over three epochs using 50 randomly selected training data samples. The loss progression in this scenario exhibits significant fluctuations in the initial stages but stabilizes in the later stages. Considering the small size of the training dataset, this early stabilization is presumed to lead to insufficient accuracy over broad data.

In finetuning 2, depicted in Fig. 4-2, the training dataset was increased to 250 randomly selected samples, and training was conducted over three epochs. Compared to the first scenario, the adaptation to a more diverse training dataset results in greater and more prolonged fluctuations. However, it is inferred that this leads to more advanced learning.

In finetuning 3, shown in Fig. 4-3, 250 training data samples were used, but the model was trained over nine epochs. In this case, the loss stabilizes within one epoch, suggesting that increasing the number of epochs does not necessarily aid in learning more from the data.

From these observations, it is evident that the size of the training data and the number of epochs have a significant impact on the learning process.

Figure 5 compares the classification accuracy of these three fine-tuned GPT-3 Turbo models and GPT-4 variants.

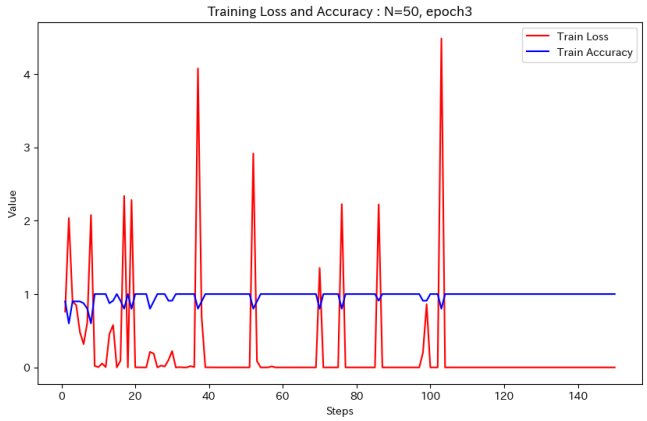


Figure 4 - 1 Training Loss and Accuracy over Epochs: N=50, Epoch=3 (red: train loss, blue: train Accuracy)

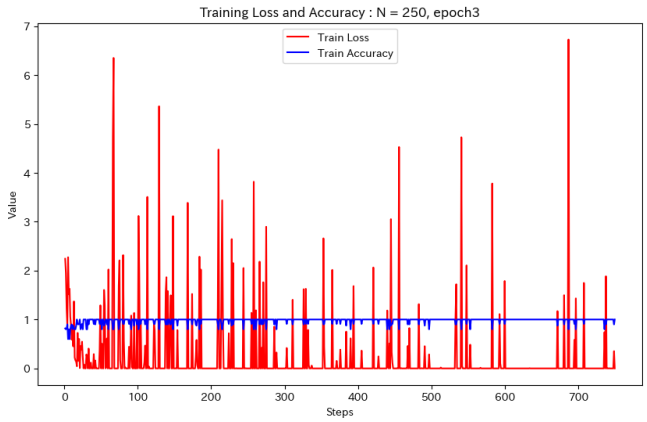


Figure 4 - 2 Training Loss and Accuracy over Epochs: N=250, Epoch=3 (red: train loss, blue: train Accuracy)

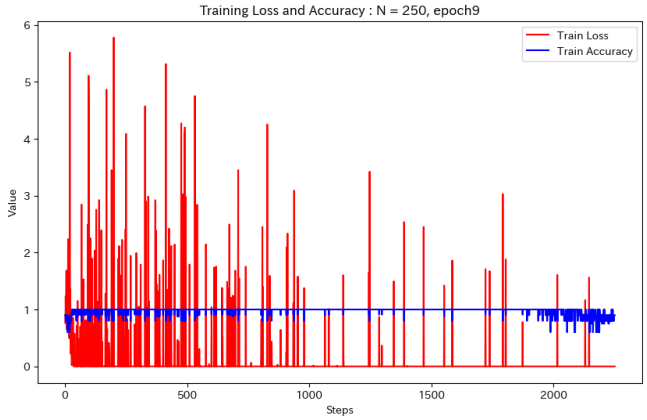


Figure 4 - 3 Training Loss and Accuracy over Epochs: N=250, Epoch=9 (red: train loss, blue: train Accuracy)

Among the metrics of Accuracy, Precision, Recall, and F1-score, we specifically chose the F1-score due to its balanced representation of both Precision and Recall. In both scenarios, where the training data consists of either 50 or 250 randomly selected samples, fine-tuning significantly improves the F1 score, reaching levels comparable to the accuracy achieved using Function Calling in GPT-4 variants. Increasing the training data from 50 to 250 samples results in a slight improvement in the F1 score. However, no additional improvements were observed with an increase in the number of epochs.

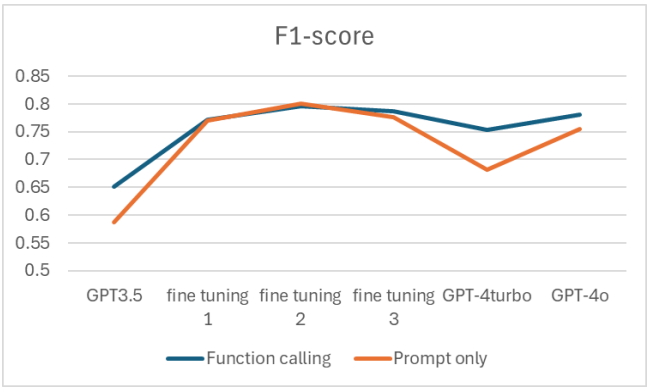


Figure 5 Comparison of F1-Scores for Function Calling and Prompt-Only Approaches Across Different Models and Fine-Tuning Stages

While using Function Calling without fine-tuning is effective, combining it with fine-tuning achieves higher classification accuracy. Furthermore, fine-tuning reduces the performance gap between using Function Calling and using only prompts. These findings provide valuable insights for selecting the optimal model design and training strategy for complex tasks such as non-functional requirements classification.

5.3 Improve Context Understanding

This experiment evaluated the effectiveness of the proposed sentence combination method by comparing it to analyses conducted without sentence concatenation. The classification accuracy results for each metric are shown in Table 5.

As illustrated in the table, the classification method using sentence combination (Combined Analysis) outperforms the no concatenation analysis across all evaluation metrics. Specifically, the accuracy improved.

Out of 414 experimental data points, 22 data points that were previously incorrect were corrected by using sentence combination, whereas 6 data points that were previously correct were misclassified. These results indicate that incorporating contextual information enhances classification accuracy, although it also introduces new misclassifications.

Basic no concatenation analysis methods tend to overlook contextual information as they consider only the target sentences for classification. In contrast, the method using sentence combination adds contextual information by combining preceding and/or succeeding sentences, allowing for more accurate classification. This approach is particularly effective for documents like requirement specifications, where non-functional requirements often span multiple sentences.

Table 5 Classification Accuracy Results Comparing the Proposed Sentence Combination Method and no Concatenation Analysis

	without Sentence Concatenation	Combined Analysis
Accuracy	0.792	0.831
Precision	0.785	0.839
Recall	0.779	0.808
F1-score	0.780	0.818

From these results, it can be confirmed that the classification method using sentence combination is effective, especially in the classification of non-functional requirements, by leveraging contextual information to improve classification accuracy. However, the increase in misclassifications suggests that further improvements and optimizations are necessary.

6 DISCUSSION

The primary aim of this study was to demonstrate that incorporating the Function Calling feature into ChatGPT's classification process can enhance the accuracy of NFR classification compared to relying solely on prompts. By adding Function Calling to the same prompt, the method achieves improved accuracy by reducing ambiguities and ensuring adherence to predefined classification criteria, even in a zero-shot setting.

This improvement, as indicated by accuracy increases across all models tested, is primarily attributed to the structured data retrieval and contextually appropriate categorization enabled by Function Calling. For example, GPT-4 Turbo demonstrated an improvement in F1 score, increasing from 0.681 to 0.753. Similarly, GPT-3.5 Turbo improved from 0.587 to 0.651, while GPT-4o achieved a gain from 0.754 to 0.780.

These results highlight that the integration of Function Calling consistently enhances classification accuracy across models by ensuring adherence to predefined classification boundaries. By defining enumerated categories, the model effectively avoids hallucinations, ensuring that the classifications adhere strictly to the intended boundaries. This capability highlights the model’s alignment with human reasoning in complex classification tasks.

Furthermore, the proposed method is practical as a primary classification tool, particularly when secondary human review is integrated into the process. This approach offers high efficiency and reliability. Additionally, considering that even human reviewers may have differing opinions on certain classification targets or that some items inherently span two categories, the practicality of this method becomes even more evident. However, in critical projects, robust secondary reviews are indispensable, and further research and development are necessary to enhance classification accuracy.

The supplementary experiments aimed to enhance classification accuracy by fine-tuning the model and incorporating contextual information, resulting in measurable improvements in precision and recall metrics.

Fine-tuning demonstrated that even minimal adjustments on small datasets could yield notable improvements in classification accuracy. For instance, the F1 score of GPT-3.5 Turbo improved significantly from 0.651 to 0.796, highlighting the effectiveness of fine-tuning in enhancing the model's performance.

Similarly, the use of sentence concatenation to incorporate contextual information from adjacent sentences enhanced the model's understanding, leading to more precise classifications. Specifically, this approach improved accuracy from 0.792 to 0.831, demonstrating the value of leveraging contextual information in classification tasks.

However, these techniques were positioned as complementary to the primary use of Function Calling.

In conclusion, Function Calling proved to be an essential feature, enabling consistent improvements in NFR classification accuracy over prompt-only methods. The supplementary methods, while beneficial, serve primarily to augment the foundational improvements achieved through Function Calling. Future research will focus on refining these methods and exploring their broader applications.

7 SUMMARY OF CONTRIBUTION AND CONCLUTIONS

This paper proposed and demonstrated the effectiveness of an automatic classification method for NFRs using Function Calling with large language models.

The Function Calling capabilities of GPT-4 Turbo and GPT-4o achieved excellent F1 scores of 0.753 and 0.780, respectively, matching or exceeding the performance of traditional task-specific models such as CNN or BERT on datasets similar to those used in this study for NFR classification tasks, even in a zero-shot learning state, demonstrating their high capability.

However, instances of misclassification were observed, particularly in scenarios involving ambiguous or overlapping categories, indicating the need for improvements in the model's contextual understanding abilities.

Two additional approaches, namely fine-tuning and sentence concatenation, were employed to enhance classification accuracy and address contextual ambiguities.

First, GPT-3.5 Turbo, with minimal fine-tuning, achieved results comparable to GPT-4 variants, emphasizing the importance and efficiency of fine-tuning. The significance of fine-tuning is evident, as it enhances the model's contextual understanding and classification accuracy.

Notably, the performance of the prompt-only approach also improves significantly with fine-tuning, as evidenced by an F1 score increase from 0.587 to 0.800 at maximum, ultimately narrowing the performance gap between the two methods.

Furthermore, experiments involving the concatenation of related sentences, by linking the sentences preceding and following the target for classification, confirmed the effectiveness of this approach in improving contextual understanding.

Future research should address these challenges by focusing on enhancing the model's contextual understanding

and creating targeted training datasets that emphasize the model's most challenging classification categories of NFRs.

ACKNOWLEDGEMENT

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

REFERENCES

[1] Japan Information Service Industry Association REBOK Planning WG, Requirements Engineering Body of Knowledge, Kindaikagaku-sha, pp. 1-200 (2011).
[2] Ministry of Economy, Trade, and Industry, Guideline for Improving the Reliability of Information Systems, 2nd Edition, pp. 1-50 (2009).
[3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, and I. Polosukhin, "Attention is all you need," Advances in Neural Information Processing Systems, Vol. 30, pp. 6000-6010 (2017).
[4] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," NAACL, Vol. 1, No. 1, pp. 4171-4186 (2019).
[5] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," OpenAI, pp. 1-10 (2018).
[6] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," OpenAI, pp. 1-15 (2019).
[7] K. Mukaida, S. Fukui, T. Nagaoka, T. Kitagawa, S. Ogata, and K. Okano, "Efficient automatic classification of non-functional requirements in information systems using deep learning," IEICE Technical Report, Vol. 123, pp. 13-18 (2023).
[8] Information-Technology Promotion Agency, Japan, System Infrastructure Non-Functional Requirements Related Grade Table, Apr. 2013.
[9] T. Kinoshita, T. Omori, and J. Onishi, "Extraction and validation of non-functional requirements from Japanese requirements documents," IPSJ Technical Report (SE), Vol. 2021.15, pp. 1-6 (2021).
[10] R. K. Gnanasekaran, S. Chakraborty, J. Dehlinger, and L. Deng, "Using Recurrent Neural Networks for Classification of Natural Language-based Non-functional Requirements," Proc. of the 4th Workshop on Natural Language Processing for Requirements Engineering, CEUR Workshop, Vol. 2857, pp. 1-10 (2021).
[11] T. Kitagawa, T. Nagaoka, "Proposal and evaluation of an automatic classification method for requirement specifications using deep learning," IPSJ Transactions, Vol. 61, No. 4, pp. 842-852 (2020).
[12] Y. Zhu, Y. Wang, J. Qiang, and X. Wu, "Prompt-Learning for Short Text Classification," IEEE Trans. Knowledge and Data Engineering, Vol. 36, No. 10, pp. 5328–5339 (2024).
[13] J. Liu and L. Yang, "Knowledge-Enhanced Prompt Learning for Few-Shot Text Classification," Big Data and Cognitive Computing, Vol. 8, No. 4, Article 43, pp. 1-10 (2024).
[14] B. J. Gutierrez, N. McNeal, C. Washington, Y. Chen, L. Li, H. Sun, and Y. Su, "Thinking about GPT-3 In-Context Learning for Biomedical IE? Think Again," Findings of the Association for Computational Linguistics: EMNLP, pp. 4497-4512 (2022).
[15] X. Sun, X. Li, J. Li, F. Wu, S. Guo, T. Zhang, and G. Wang, "Text classification via large language models," Findings of the Association for Computational Linguistics: EMNLP, pp. 8990-9005 (2023).
[16] S. Ibe, S. Kurata, T. Nagaoka, A. Furuhata, K. Goto, S. Fukui, and T. Kitagawa, "Automatic classification of requirement specifications using large language models," Proc. of the Software Engineering Symposium 2023, pp. 86-92 (2023).
[17] Y. Tan, D. Min, Y. Li, W. Li, N. Hu, Y. Chen, and G. Qi, "Can ChatGPT Replace Traditional KBQA Models? An In-Depth Analysis of the Question Answering Performance of the GPT LLM Family," Proc. of the International Semantic Web Conf., pp. 348-367 (2023).
[18] T. Brown, et al., "Language Models are Few-Shot Learners," Advances in Neural Information Processing Systems, Vol. 33, pp. 1877-1901 (2020).
[19] Q. Dong, S. Jiang, and W. Liu, "A survey for in-context learning," arXiv preprint arXiv:2301.00234, pp. 1-20 (2022).
[20] S. Min, M. Lewis, H. Hajishirzi, and L. Zettlemoyer, "Rethinking the Role of Demonstrations: What Makes In-Context Learning Work?," Proc. of the Conf. on EMNLP, pp. 11048-11064 (2022).
[21] Z. Wan, J. Xu, and M. Huang, "GPT-RE: In-context learning for relation extraction using large language models," arXiv preprint:2305.02105, pp. 1-10 (2023).

(Received: November 14, 2024)
(Accepted: August 12, 2025)

Kazuhiro Mukaida received his ME degree from Shinshu University in 2024 and is currently pursuing his doctoral studies at the same institution. His research focuses on natural language processing using large language models (LLMs) with the goal of enhancing the efficiency of requirements engineering processes.

Seiji Fukui received his M.S. in Natural Science from Okayama University in 2019. He is currently working for Toshiba Corporation. His research interests include software development environments and the application of AI technologies to software engineering.

Takeshi Nagaoka received his M.I. and Ph.D. degrees in Information Science and Technology from Osaka University in 2008 and 2011, respectively. He is currently working for Toshiba Corporation. His research interests include formal methods, software development environments, and the application of AI technologies to software engineering.

Takayuki Kitagawa received his M.E. degree in Management and Information Science from the Prefectural University of Hiroshima in 2008. He is currently working for Toshiba Corporation. His research interests include requirements engineering, software lifecycle process, and the application of AI technologies to software engineering.

Shinpei Ogata is an Associate Professor at Shinshu University, Japan. He received his B.E., M.E., and Ph.D. from Shibaura Institute of Technology in 2007, 2009, and 2012, respectively. He served as an Assistant Professor at Shinshu University from 2012 to 2020 and has been an Associate Professor there since 2020. He is a member of IEEE, ACM, IEICE, IPSJ, and JSSST. His current research interests include model-driven engineering for information system development.

Kozo Okano received his BE, ME, and PhD degrees in Information and Computer Sciences from Osaka University in 1990, 1992, and 1995, respectively. He was an Assistant Professor and an Associate Professor of Osaka University. In 2002 and 2003, he was a visiting researcher at the Department of Computer Science of the University of Kent in Canterbury, and a visiting lecturer at the School of Computer Science of the University of Birmingham, respectively. Since 2020, he has been a Professor at the Department of Electrical and Computer Engineering, Shinshu University. Since 2023, he has been the Director of Center for Data Science and Artificial Intelligence. His current research interests include formal methods for software and information system design and applying deep learning to Software Engineering.He is a member of IEEE, IEICE, and IPSJ.

Regular Paper

A Study on Power Distribution Method Using Electric Vehicles

Masashi Saito*, Wataru Kunimatsu*, Hikaru Akutsu**, Chiaki Kojima**,
Kazutoshi Sakakibara†, Hironao Kawamura‡

* Graduate School of Engineering, Kanazawa Institute of Technology, Japan
** Department of Electrical and Electronic Engineering, Toyama Prefectural University, Japan
† Department of Data Science, Toyama Prefectural University, Japan
‡ Innovation Laboratory, Hokuriku Electric Power Company
msaito@neptune.kanazawa-it.ac.jp, b1846432@planet.kanazawa-it.ac.jp
hikaru_akutsu@pu-toyama.ac.jp, chiaki@pu-toyama.ac.jp, sakakibara@pu-toyama.ac.jp
h.kawamura@rikuden.co.jp

Abstract - We have developed a simulator to establish energy management algorithms for realizing power distribution over wide areas using electric vehicles. This simulator is for a service called a virtual power distribution network, which not only eliminates waste power from solar power generation, but also cuts peak power demand.

We have set up the simulation with the area being 20km square, with roads running north-south and east-west every 1km. This area is divided into three categories: commercial, industrial, and residential. We have evaluated and verified changes in power generation and demand throughout the day by setting use cases for power generation, demand, and power storage systems and simulating power transportation using thousands of electric vehicles.

Simulation results show the amount of electricity purchased from the grid in the industrial area could be reduced by 92%, by using power transmission part time job use case (EV part-timer), which transmits power actively using price differences in buying and selling. Using this mode, power distribution becomes very active, but this requires much more charge/discharge points. The number of required charge/discharge points increase squaring.

In this simulation, only data from the Hokuriku region is used for electricity consumption and solar power generation. By replacing these data, the feasibility of power distribution can be verified for all regions, showing that a wide variety of simulations in social systems is possible.

Keywords: Power Distribution, EV, Wide Area Simulation

1 INTRODUCTION

The introduction of renewable energy is accelerating toward a decarbonized society, and photovoltaic power generation (PV) is attracting attention. However, due to the shutdown of some power plants due to natural disasters such as earthquakes, and the increased use of air conditioning equipment due to falling temperatures, electricity demand has increased, and it is becoming an issue that daily peak demand for electricity cannot be met [1] [2].

On the other hand, the amount of electricity generated by renewable energy in Japan increased about four times in fiscal 2020 compared to fiscal 2012, resulting in a surplus of electricity supply, and output control from April to September 2023 reached 194 times, about three times the previous year. In areas with high electricity demand, such as the Tokyo metropolitan area, requests to save electricity were made, and there were concerns about a shortage of power supply. Such unstable supply and demand of electricity could lead to power outages.

In addition, there is a situation where new construction of power transmission network facilities is not progressing due to opposition from surrounding areas, and measures to deal with the aging of existing power transmission networks are also a heavy burden [3] (Fig. 1).

The Japanese government plans to increase the proportion of renewable energy to 36-38% in FY2030, which is about double the proportion in FY2019, and has estimated that 6 trillion to 7 trillion yen will be needed to develop the power grid in 2022 [4]. Therefore, it can be said that conventional power generation and transmission capacity is not enough to meet the demand.

Sustainable and stable supply of energy, including electricity, as well as decarbonization and low carbonization are important issues, and from this perspective, the introduction of renewable energy in the power sector and the introduction of EVs (electric vehicles) in the transportation sector are expanding. EVs are equipped with large-capacity storage batteries as a power source, so they are expected to be used not only as mobile vehicles but also in the power sector. Assuming that the penetration rate of EVs increases to 50% of new car sales by 2050, it is estimated that the power

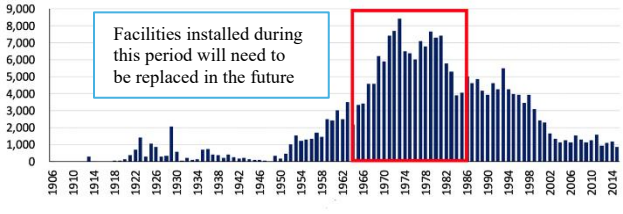


Figure 1: Distribution of Existing Steel Towers by Manufacturing Year in 2015

consumption used by EVs at that time will increase to 46.3 billion kWh, up 164.3% from 2016 [5].

EVs are equipped with large-capacity storage batteries as a power source, so they are expected to be used not only as mobile vehicles but also in the power sector. Representative examples include (i) providing energy services to the power grid and consumers by connecting EVs to the grid and charging/discharging their storage batteries, and (ii) serving as an emergency power source during disasters or grid blackouts.

In this research, to realize the role of (i) above, we have developed a simulator that uses EVs to transport electricity to places where there is a power shortage. Using the developed simulator, we report the results of evaluating and verifying the simulation results to see whether a power transmission system can be established without changing the current grid system.

2 RELATED WORKS

Ichii et al. [6] examined the possibility of EVs replacing privately operated power lines as a means of making up for power shortages in smart towns, which is an important issue in the use of renewable energy. As a result, the possibility of EVs replacing privately operated power lines was low in satellite smart towns (SSTs) close to distributed power sources, but in small, remote SSTs, the transportation costs of EVs were lower than those of privately operated power lines, indicating the possibility of EVs replacing privately operated power lines. In addition, with improvements in electricity self-sufficiency and advances in DC power supply technology, it is expected that costs will be reduced by using EVs. It was also suggested that there is a high possibility of low-cost use of EVs for electricity transportation. This suggests that costs can be reduced for SSTs by having EVs take on the role of electricity transportation, and that transmission separation could be an opportunity to promote the use of renewable energy.

Tanoue et al. [7] aimed to build a support system for power transportation by EVs. Although the power consumption characteristics of EVs vary depending on the driving state, distance, and altitude; to facilitate the use of geographic information systems, they proposed a method to estimate the power consumption of EVs while driving by separately expressing the power consumption of distance and altitude difference. As a result, they derived an equation for estimating the power consumption of EVs while driving using geographic information. They also showed that it is possible to visually grasp the EV power by visualizing it using the derived estimation equation.

Urabe et al. [8] designed a new EV infrastructure system that realizes peak shift of power by providing power (batteries) to users within the power supply area using EVs and considered the leveling of power demand. As a result, they showed the effectiveness of efficiently using the batteries provided to users from three perspectives: the optimal number of EV infrastructure systems to be placed within the area, the optimal facility placement of EV infrastructure systems within the area, and a plan for providing power to users within the area using the EV infrastructure system.

Murakami [9] used agent-based simulation to examine the problem of power load concentration due to the spread of electric vehicles (EVs). He showed that random charging start times can sometimes achieve high power load leveling effects. He showed that autonomous information exchange and interactions among EV agents can achieve higher power load leveling. He focused on information exchange within small groups and suggested that power load leveling can be achieved even in large-scale systems without information exchange between independent groups.

Takagi et al. [10] focused on autonomous decentralized control that did not require additional investment in communication infrastructure and proposed two measures to mitigate the impact on the grid that reflected the distribution characteristics of daily driving distances and ensured convenience for EV users. As a result of the evaluation, it was found that the magnitude of the steep peak was proportional to the number of EVs charging at the same time, so along with the EV penetration rate, how frequently EV users charge was also an important factor. However, since the frequency of charging depended on various uncertain factors such as the battery capacity of EVs that would become more widespread in the future and the specifications of chargers, it was shown that even under the conditions assumed in this paper where charging impacts were most likely to occur, it was possible to mitigate the peak due to EV charging load by adopting a quadratic function case, etc.

Moriya et al. [11] proposed a method of power trading using virtual power plants in smart grids and the use of PHEVs and EVs as storage batteries to reduce peak power demand across the entire power grid, but they did not consider the driving of PHEVs and EVs. They also proposed a method of power leveling that considered the driving of PHEVs and EVs. As a result, when comparing timer charging and bottom charging, it was found that timer charging had a greater leveling effect when there were many EVs, and they showed that the power load could be reduced by using EVs.

Mustapha Aachiq et al. [12] considered electric vehicles (EVs) as a means of adjusting demand in homes and created a model to simulate what kind of EV battery operation would be optimal for consumers when linking EV batteries with PV systems under the electricity pricing system that will become more diverse in the future. The results showed that the impact of FIT prices on EV battery operation and the economics of each household is that when the FIT PV purchase price falls, the amount of reverse power flow is suppressed and the amount of charging from PV to EV increases. In addition, it is suggested that V2G can contribute to load leveling in the grid if a large number of PHEVs are introduced and electricity prices in homes are determined in conjunction with the marginal fuel cost of the grid.

These studies showed the possibility of transmitting power over a small area and meeting electricity demand.

In small-scale areas, it is easy to link EVs and power storage systems, and it is possible to respond to peak power demand even after power demand has increased. However, in medium-scale or larger areas, it is difficult to link EVs and energy storage systems, so it is necessary to verify whether it

is possible to meet the power demand in areas with a power shortage through verification. Therefore, verification is carried out in a medium-scale area assuming an increased number of vehicles and a map of 20 km x 20 km.

In this study, we have developed a simulator based on various use cases in which EVs are charged using surplus solar power generation electricity, and in areas where there is a power shortage, EVs transport power to discharge it, and the effective use of surplus power from power transport and the EV infrastructure required to respond to peak power demand are verified. In addition, three patterns of power demand are prepared: commercial, industrial, and residential areas [13].

3 OVERVIEW OF ELECTRIC POWER TRANSMISSION USING EVs

3.1 Electric Power Transmission by EV

We have set up various use cases to realize a charging/discharging service using EVs in medium-sized area energy management and verify it using a simulator to investigate the impact on the supply and demand balance. Specifically, we have built a system in which surplus electricity generated by PV is charged into EVs that stop at charging/discharging points, and then transmitted to other charging/discharging points for discharging. An overview of power transmission is shown in Fig. 2.

3.2 Overview of Our EV Energy Management System Simulator

The EV energy management system is positioned between existing energy management (power control) systems and MaaS (Mobility as a Service) services, which will continue to develop in the future, and proposes charging and discharging services to moving EVs through road network. Figure 3 shows the overview of the EV energy management system.

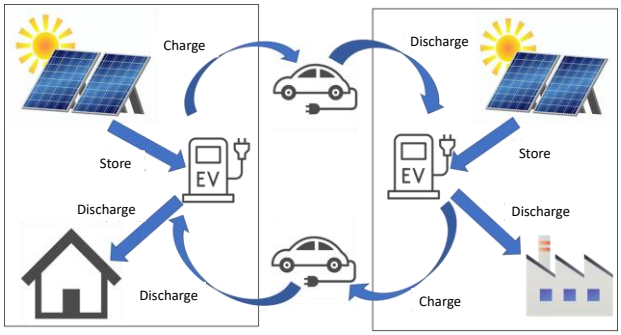


Figure 2: An overview of electricity transportation

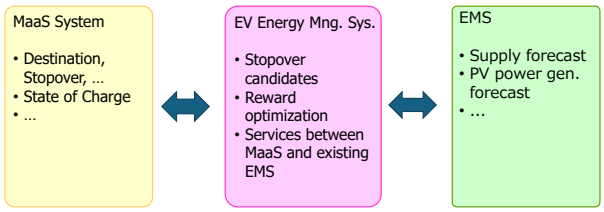


Figure 3: EV Energy Management System Overview

Unfortunately, there is no MaaS system that can be freely used in an EV energy management system, and the energy management system cannot be used. Therefore, we have developed the "EV Energy Management System Simulator (hereinafter EVEMSS)" to simulate electricity transportation. EVEMSS is composed of a vehicle information simulator instead of a MaaS system, an EV energy management system, and an existing energy management system that provides data on the grid DB that manages the power demand and PV power generation managed by the energy management system (Fig. 4). The reason why the Power Grid is connected to charging/discharging points is because the surrounding area's electricity is provided via the charging/discharging point.

By using EVEMSS to investigate and analyze the impact in various use cases, it is possible to use it as a basic tool for verifying and proposing new power supply systems and their services.

The EV energy management system is a subsystem that realizes customer value by customers who drive EVs stopping at places called charging/discharging points. This subsystem provides a buying and selling intermediation service by effectively utilizing surplus electricity generated by solar power generation, realizing a win-win relationship in which EV owners can purchase cheap electricity and PV owners can sell waste electricity that they are forced to discard. Through such intermediation, we hope to determine the possibility of creating new service businesses.

3.3 EVEMSS System Configuration

The system configuration of the EVEMSS is shown in Fig. 5. The vehicle information simulator drives the vehicle using

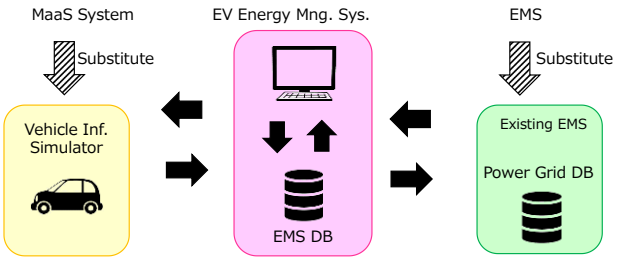


Figure 4: EV Energy Mng. System Simulator Overview

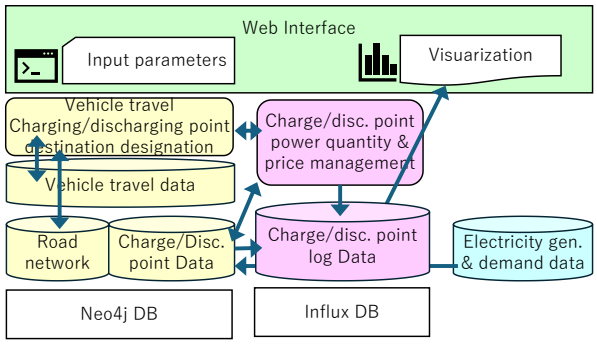


Figure 5: EVEMSS System Configuration

data for vehicle movement, road networks, and charging/discharging point data. When charging/discharging becomes necessary, it queries the EV energy management system to determine the route and destination. The EV energy management system manages the amount of electricity and buying/selling prices at charging/discharging points, as well as managing the power logs of the charging/discharging points. The existing energy management system stores data on the amount of electricity generated and demand at each location.

3.4 Road Network in EVEMSS

The system has been currently built assuming a Manhattan model road network. A tertiary mesh code is used for this purpose. The distance of a tertiary mesh code is not constant because the Earth is a sphere, but in the Hokuriku region it is 926 m north to south and 1,118 m east to west. EV operation is simulated on a 20 x 20 grid road network divided into tertiary meshes [14].

The roads defined above have a graph (network) structure. In order to make it possible to represent any road in any location, we are using Neo4j, a NoSQL database that makes it easy to model graph structures.

We assume that the data on charging and discharging points is held by the MaaS service. Each node stores the amount of battery power held in the battery in that area, the price at the time of buying and selling and so on.

3.5 Visualization of EVEMSS

In order to simplify the execution of EVEMSS and the evaluation of the simulation results, a web-based front end has been developed, which allows the selection of input data files, the display of the overall simulation results, and the detailed examination of individual results interactively. This makes it possible to execute simulations of multiple scenarios by simply replacing the data in EVEMSS.

Figure 6 shows an example of charging/discharging point power amount.

Vehicle information is set one by one, including the vehicle identifier, departure time, charging/discharging behavior, initial battery value, departure point, multiple intermediate destinations, and destination.

Electricity generation and demand are set every 30 minutes.

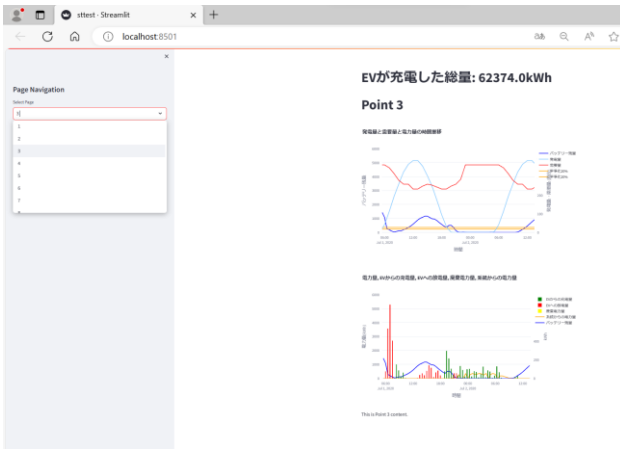


Figure 6: Visualization of EVEMSS

4 SIMULATION SETTINGS

4.1 Road Network and Charging/ Discharging Point Settings

We assume the Manhattan model for the road network as explained in Section 3.4. In addition, we arrange the charging /discharging points as power demand locations, with commercial areas, industrial areas, and residential areas. We also set up 16 charging/discharging points as representative locations of 5km square at the center of the area (Fig. 7).

Each charging/discharging point has 900 kWh power batteries, which are used as a charging and discharging infrastructure in each region as a grid network. In addition, like a typical gas station, multiple-bidirectional chargers are set up to be available, and in this study, up to six units can be traded at one charging/discharging point.

4.2 Power Generation and Demand Setting

Regarding the use case of the power generation, demand, and storage system, the season of power generation is set to one day in summer. The day with the most solar radiation in the year is set based on data from the Japan Meteorological Agency for Kanazawa City [15], and the simulation has been performed based on that.

Power demand is created based on the power load level of the Institute of Electrical Engineers of Japan model. The specific values are created originally based on data from the Architectural Institute of Japan [16].

The amount of power generated by PV and the amount of power demand in each area are shown in Fig. 8. Simulations have been performed based on these values from now on. The graph shows the amount of power generated in blue, the power demand in residential areas in orange, the industrial

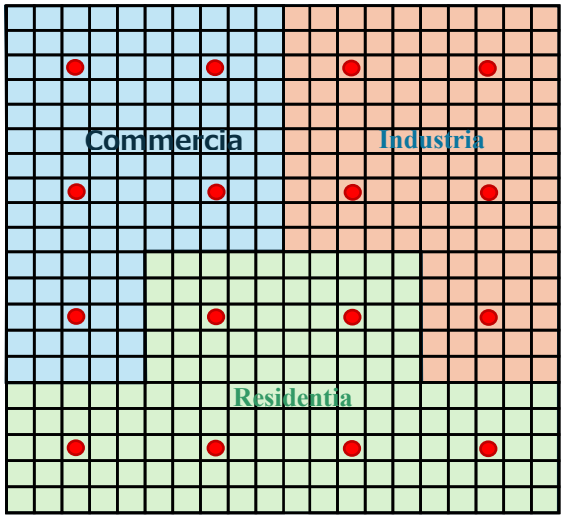


Figure 7: Road Network and Charging/Discharging Points in EVEMSS

areas in gray, and the commercial areas in yellow. In addition, the time setting is set to 5:00 a.m. to 2:00 p.m. the following day to take into account the effects of the late-night hours when power consumption is high.

4.3 Vehicle Driving and Behavior Settings

Vehicle use cases are set using data from the Japan Road Traffic Information Center (JARTIC) [17]. The settings are based on the traffic volume on Route 8 at Okyozuka in Ishikawa Prefecture, and every 10minutes in the EVEMSS the number of the traffic volume EVs start driving.

Each vehicle runs two laps of approximately 80km circumference at a speed of 40km/h. The maximum battery capacity of the EV is set at 70kWh, each EV start searching charging/discharging point at 20% of battery capacity for charging and at 95% for discharging. A discharging EV stops discharging at 20% of battery capacity.

Two patterns of EV behavioral changes are prepared for the vehicle's behavior settings. The first is a vehicle that prioritizes distance when heading to a charging/discharging point. The second is a EV part-timer mode. A distinctive feature of this vehicle's behavior settings is that it does not sell power unless the selling price when discharging is at least 10 yen higher than the selling price when charging.

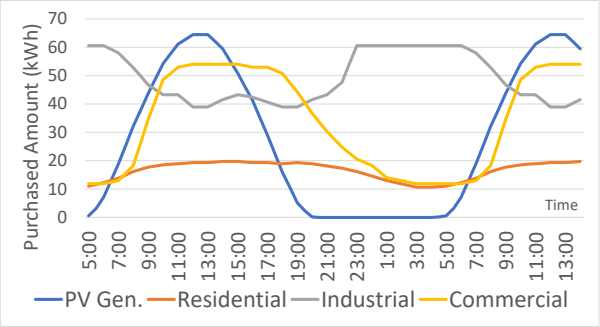


Figure 8: The Amount of Power Gen. by PV and the Amount of Power Demand of Each Area

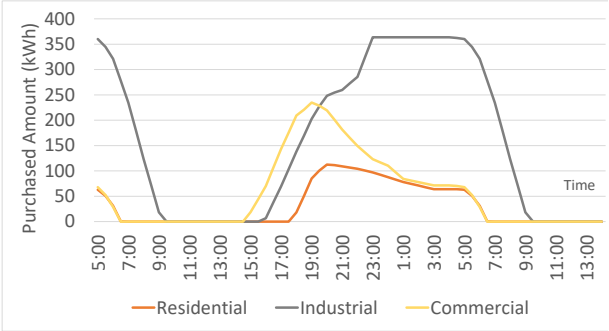


Figure 9: Use of PV Power Generation

5 SIMULATION RESULTS

5.1 Use of PV Power Generation and Battery

Prior to conducting the simulation, electricity is purchased from the grid by time based only on the amount of electricity generated by PVs, and the results are shown in Fig. 9.

Commercial areas need to purchase 3,852 kWh of electricity from the grid, industrial areas 10,822 kWh, and residential areas 2,906 kWh.

The results in Fig. 9 show that the amount of electricity purchased from the grid is concentrated from the evening through the night, and it can be said that the bias in the time periods for electricity purchases makes it impossible to stabilize the electricity supply and demand balance.

Next, we simulate with PV power generation and batteries. Simulation results are shown in Fig. 10. The presence of the battery is enough to cover the electricity needs of the residential area. However, the industrial area needed to purchase 7,572 kWh of electricity from the grid, and the commercial area needed to purchase 1,202 kWh. The amount of electricity purchased from the grid in the industrial area is reduced by approximately 30% compared to when there is no battery. As for the residential area, the result shows that surplus electricity is generated. It can also be seen that even with the use of the power storage system, the amount of electricity purchased during the nighttime hours is high. Therefore, it is necessary to consider measures to ensure electricity for commercial and industrial areas at late-night.

5.2 Evaluation of Electric Power Transmission by EVs

5.2.1 Evaluation Assuming EVs Account for 2.5% of Traffic Volume

The evaluation is carried out assuming that 2.5% (629 vehicles) of vehicle traffic are EVs, 50% of the EV departures are from drivers who prioritized distance, and the remaining 50% are from drivers who are driving in EV part-timer mode.

The results in Fig. 11 show that commercial areas need to purchase 588 kWh of electricity from the grid, industrial areas 2,996 kWh, and residential areas 570 kWh. It can be seen the

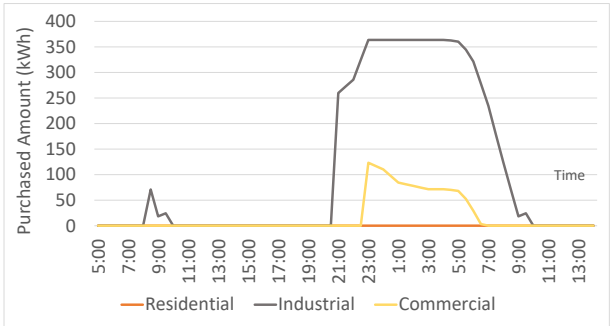


Figure 10: Use of PV Power Generation and Batteries

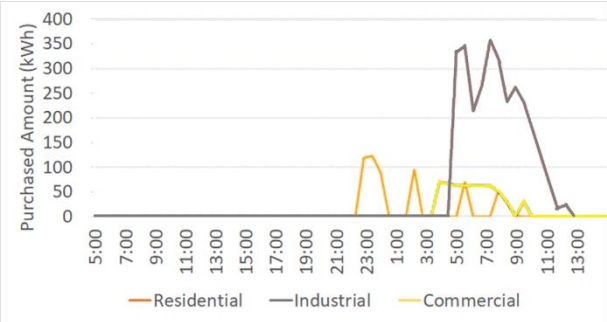


Figure 11: Use of 2.5% vehicle being EV

amount of electricity purchased from the grid in industrial areas is reduced by approximately 60% compared to when an evaluation is conducted using only the energy storage system. It also shows that the generation of surplus electricity in residential areas can be eliminated by conducting electricity transmission trading. However, it can be seen from the results that the amount of electricity purchased during late night hours has not been reduced compared to Fig.9.

5.2.2 Evaluation Assuming All 2.5% of Vehicle Traffic Volume Being in EV Part-timer Mode

The number of EVs is set at 2.5% of the vehicle traffic volume, and the evaluation is carried out assuming that all vehicles are driven by drivers using EV part-timer mode.

From the results in Fig. 12, the industrial area needs to purchase 4,551 kWh of electricity from the grid, and the residential area needs to purchase 645 kWh. It can be seen that the amount of electricity purchased from the grid in the industrial area is reduced by approximately 40% compared to when an evaluation is performed using only the energy storage system.

Effect of EV power transmission in industrial areas

We evaluate whether electricity purchases from the grid during the night and late-night hours in industrial areas can be reduced by electricity transportation and evaluated the impact of the number of EV vehicles and distance-priority mode and electricity part-timer mode shown in Table 1.

From the above results, the amount of electricity purchased from the grid in industrial areas can be reduced by using EVs for power transmission, but the effect of power transmission decreases when the number of vehicles exceeds a certain level. Therefore, a simulation is performed with the following settings:

Table 1: Amount of Electricity Purchased from the Grid Using both Modes and Part-Timer Mode

The number of Vehicles	Purchased Amount Both Modes	Purchased Amount Part-Timer Mode
629	2,996 kWh	4,551 kWh
1,246	2,085 kWh	2,585 kWh
2,492	3,722 kWh	1,403 kWh
3,738	3,482 kWh	3,607 kWh

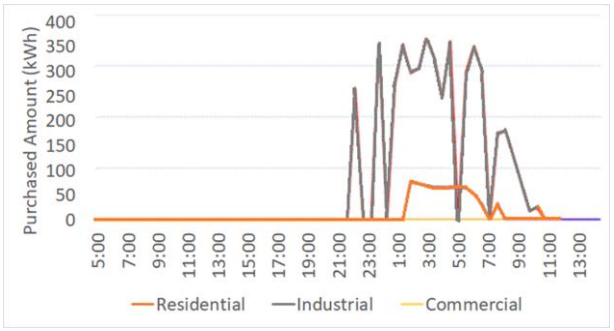


Figure 12: Use of EV Part-timer Mode

(1) When transporting electricity to industrial areas, charging is not performed in the industrial areas, only discharging is performed,
(2) The price is changed by time of day, and the behavior of the electric power part-time mode vehicles is changed.

By implementing these measures, the amount and timing of electricity purchased from the grid in industrial areas and other areas changes, but it does not necessarily work well. The application of a more sophisticated algorithm is necessary.

5.3 The Number of Vehicles and Discharge Amount

Figure 13 shows how much discharge is possible depending on the number of vehicles being charged and discharged simultaneously. Kanazawa city has approximately 80,000 vehicles, of which 4,000 vehicles are in operation if the vehicle utilization rate is 5%. If the scale of Kanazawa

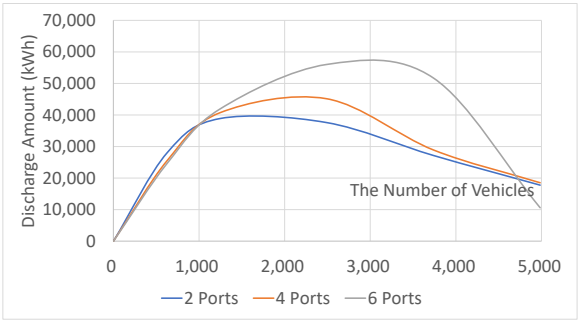


Figure 13: Relationship Between the Number of Vehicles and Discharge Amount

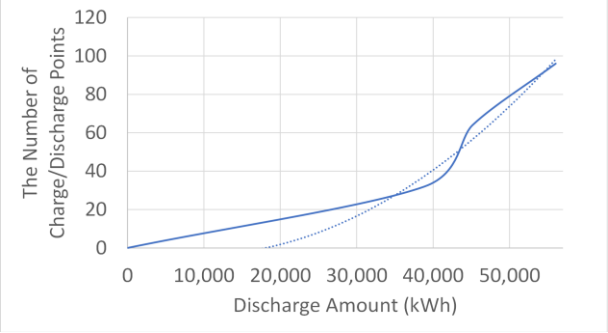


Figure 14: Required Chage/Discharge Points for Discharge Amount

city and the EV energy management system simulator are combined, the number of vehicles is approximately 3,400 vehicles if the vehicle utilization rate is 5%, so it can be said that an infrastructure environment capable of accommodating at least this number of vehicles is necessary.

Figure 14 is a graph showing the maximum discharge amount versus the number of charge/discharge points. The approximation formula is expressed as follows:

$$y = 0.00000005x^2 - 0.0008x - 0.23$$

From Fig. 12 and the approximation equation, it is possible that in order to increase the amount of power transmitted, the charge/discharge points must be increased by the square of the amount of power.

6 CONCLUDING REMARKS AND FUTURE WORKS

In this study, we set up a specific use case to realize a charging and discharging service using EVs in energy management in a medium-sized area and verified it using a simulator to investigate the impact on the supply and demand balance. It is shown that by transporting electricity using EVs, surplus electricity can be reduced to 0 kWh and peak electricity demand can be cut.

In order to solve the power shortage in the industrial area at night and late at night, we evaluate it with a setting that all EV part-timer mode and no charging from the industrial area but discharging to the industrial area. This changes the amount and timing of electricity purchased from the grid in the industrial area and other areas, but it does not necessarily work well, more sophisticated algorithm is necessary.

In addition, under the current infrastructure setting, there are not enough charging places for EVs when the number of vehicles exceeds a certain level, so it can be said that a charging and discharging infrastructure suitable for the number of vehicles is necessary.

From the relationship between the number of vehicles and the amount of discharge, it is found that in order to increase the amount of electricity transmitted, the number of vehicles trading at the charging and discharging points must be increased by about the square of the amount of power amount. In the future, when verifying by increasing the number of vehicles, it can be said that the number of charging/ discharging points must be increased at the same time.

Based on the above, future issues include power leveling and incentives for vehicles.

Regarding the power demand in industrial areas, this model is for the case where the power company promotes operation at night to prevent power shortages in industrial areas. Therefore, future verification will also need to verify the case where there is a peak in power demand during the daytime when the power company's power usage shift is not implemented. This will become increasingly important in the future when power supply by PV becomes common.

As future works, we would like to conduct simulations based on more realistic road networks, vehicle driving, charging/discharging point placement and power demand, and show that it is possible to provide power transmission by EVs in specific areas, and even services such as home delivery of power by EVs. In fact, the current EVEMSS does not scale with respect to the number of EVs. In addition, from the perspective of the power company, if it becomes possible to implement measures such as leveling the supply of power and limiting the upper limit of the amount of power transmitted in the grid, we would like to aim for power transportation by EVs to emerge as a real service.

REFERENCES

[1] Ministry of Economy , Trade and Industry Japan, "Draft report on the tight electricity supply and demand situation in eastern Japan in March 2022", pp3-7 (2022).

[2] A. Nakajima et al., "Impact of the shutdown of large-scale generators due to the Great East Japan Earthquake on JEPX electricity spot prices", Proceedings of the Tohoku Branch Joint Conference of the Institute of Electrical Engineers of Japan (2012).

[3] Agency for Natural Resources and Energy, “Current energy situation and issues”, https://www.meti.go.jp/ shingikai/enecho/denryoku_gas/denryoku_gas/pdf/045 _04_02.pdf, (accessed Jan. 25, 2023).

[4] Yomiuri Shimbun, Electricity Crisis <2> Poor power transmission, renewable energy waste, (2023/11/21).

[5] M. Fujiyama, "Electricity consumption in 2050 will decrease by 20% compared to 2016 - Population decline and progress in energy conservation will significantly push down electricity consumption", Japan Research Institute Research Focus, No.2018-003 (2018).

[6] K. Ichii et al., "How to transport electricity to satellite-type smart towns", Urban Planning Papers, No.3, Vol.55, pp569-576 (2020).

[7] K. Tanoue et al., "Estimation of Electric Vehicle Power Transport Capacity Using Geographic Information Based on Vertical Power Cost Calculation", Proceedings of Kyushu Joint Conference of Institutes of Electrical Engineers of Japan (2015).

[8] S. Urabe et al., "Optimal Design Method of EV Infrastructure System Considering Power Demand Leveling - By Providing Batteries to Users in Power Supply Areas -", Information Processing Society of Japan, No. 3, 2012-MPS-91, pp1-7 (2012).

[9] T. Murakami et al., "Agent Simulation of Power Load Leveling When Electric Vehicles are Connected to the Grid", IEEJ Transactions on Electrical Engineering, Vol. C (Electronics, Information and Systems Division), No. 9, Vol. 133, pp1658-1662 (2013).

[10] M. Takagi et al., "Nighttime Charging Load Leveling Strategy Considering User Convenience of Electric Vehicles," IEEJ Transactions on Energy and Power Systems, Vol. 11, Vol. 134, pp. 908-916 (2014).

[11] T. Moriya et al., "Research and Use of Electricity Trading Using Virtual Power Plants in Smart Grids," Proceedings of the 75th National Conference, Vol. 1, Vol. 201, pp. 783-784 (2013).

[12] M. Aachiq et al., "Evaluation of Residential PV-EV Systems for Supply-Demand Balance of Power Systems," IEEJ Transactions on Energy and Power Systems, Vol. 1, Vol. 135, pp. 27-34 (2015).

[13] The Institute of Electrical Engineers of Japan, "Regional Supply System Model," https://www.iee.jp/

pes/ele_systems/base_model/overview/, (accessed Sep. 1, 2023).
[14] Ministry of Land, Infrastructure, Transport and Tourism, "Third Mesh Land Use Data - National Land Numerical Information", https://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-L03-a.html, (accessed Jan. 25, 2024).
[15] Japan Meteorological Agency, "Search for Past Weather Data", https://www.data.jma.go.jp/stats/etrn/index.php, (accessed Sep. 1, 2023).
[16] Architectural Institute of Japan, Research Committee on Energy Consumption in Residential Buildings, "Energy Consumption Database in Residential Buildings", http://tkkankyo.eng.niigata-u.ac.jp/HP/HP/database/index.htm (accessed Jul. 1, 2023).
[17] JARTIC, "Japan Road Traffic Information Center: JARTIC", https://www.jartic.or.jp/, (accessed Sep. 1, 2023).

(Received November 14, 2024)
(Accepted August 12, 2025)

Masashi Saito received B.E. degree from Tokyo Institute of Technology in 1983, M.E. degree from Cornell University in 1992 and Ph.D. degree from Osaka University in 2006. In 1983, he joined Mitsubishi Electric Corporation and has developed engineering workstations, Internet TVs, cellular phones, car navigation systems especially for operating system extension, Internet services and distributed processing. In 2006, he moved to Mitsubishi Electric Research Laboratories as a senior principal technical staff. Since 2015, he has been a professor at College of Informatics and Human Communication, Kanazawa Institute of Technology. His current research interests include ITS, distributed systems, smart communities and smart social systems. He is a member of IEEE and IPSJ

Wataru Kunimatsu received Bachelor of Engineering from Kanazawa Institute of Technology in 2022, Master of Engineering from Graduate School of Engineering, Kanazawa Institute of Technology. He joined Hagiwara Electric Holdings Co., LTD. in April 2025.

Hikaru Akutsu received the Ph.D. degree in engineering from Nagaoka University of Technology in 2020 and joined DAIHEN Corporation in the same year. In 2023, he became an assistant professor in the Faculty of Engineering at Toyama Prefectural University. He is engaged in research on the control of distributed power supply and demand networks.

Chiaki Kojima received Ph.D. degree in informatics from Kyoto University in 2007. In the same year, he joined the Department of Information Physics and Computing, the University of Tokyo as an Assistant Professor. In 2017, he joined the Department of Electrical and Computer Engineering, Toyama Prefectural University as a Lecturer. From 2020, he is currently an Associate Professor of the Department of Electrical and Electronic Engineering of the same university. He is engaged in research on system control theory and its applications to power and energy. He is a member of the Society of Instrument and Control Engineers, the Institute of Systems, Control and Information Engineers, the Institute of Electrical Engineers of Japan, and the IEEE CSS/PES/CASS.

Kazutoshi Sakakibara received his PhD in Engineering from Kobe University in 2004. He was an assistant professor from 2004 to 2008, a lecturer from 2008 to 2013 at Ritsumeikan University, and an associate professor from 2013 to 2024 at Toyama Prefectural University. He is a professor at Toyama Prefectural University. His present research interests include system modeling, emergent systems, and the development and application of optimization and machine learning algorithms. He is a member of SICE, ISCIE, SSJ, ISIJ, JSBA, and IEEJ in Japan.

Hironao Kawamura graduated from the Department of Electrical and Information Engineering, Faculty of Engineering, Kanazawa University in March 1998. He joined Hokuriku Electric Power Company in April of the same year and has been engaged mainly in the study of lightning protection design on transmission systems. He is currently affiliated with the company's Innovation Laboratory. He focuses on the planning of research and development. He holds a Ph.D. in Engineering.

Practical Paper

Audio Signal Compression and State Discrimination in a Surround Environment Using Wavelet Transform

Miyuki Shirai*, Yuhi Shuno*, Hiroki Yamamoto*, Sho Ishikawa*, and Mikiko Sode*

*National Institute of Technology (KOSEN), Niihama College, Japan
{m.shirai, m.sode}@niihama-nct.ac.jp

Abstract - When a machine makes an anomalous noise, it is often necessary to take measures such as stopping the factory lines. Thus, we have been making a system to detect machine failure using sounds. The feature of the proposed system is that it converts voice data into an image using wavelet transform, and then uses the image as input to determine anomalies using machine learning. The important thing in this system is the size of the audio data to make it easier to transmit. In addition, the evaluation criteria are based on the learning accuracy of machine learning, rather than the traditional judgment by human ears. In this paper, audio signal compression using wavelet transform is discussed. We will consider reducing the size of audio data without removing abnormal sounds contained in the audio data. The features of abnormal sound on the time-frequency plane by applying some different conversion methods are compared each other. Using images showing the obtained features, machine learning was used to distinguish between normal audio and audio containing abnormalities.

Keywords: Wavelet, Features extract with sound processing.

1 INTRODUCTION

When an abnormality occurs in a machine operating in a factory, it is necessary to stop the factory line in order to deal with the problem. However, this reduces the factory's operating hours, and if the abnormality goes unnoticed for a long period of time, even if the factory line is stopped and measures are taken, it may result in significant losses for the company. Therefore, it is necessary to detect abnormalities or signs related to abnormalities and respond to them quickly.

Factory equipment always makes some kind of noise, and if an abnormality occurs, the machines make sounds that are different from normal sounds. Only experienced engineers can naturally detect the slightest anomalous machine noises, and up until now, maintenance and inspection of equipment at factory production sites has relied on the experience and intuition of veteran engineers. In recent years, due to the influence of generational change in companies, there has been an issue regarding the succession of maintenance work techniques at factories and other work sites. Thus, factories are becoming increasingly smart factories. This technology detects failures and signs of failure in mechanical equipment without relying on human intuition.

In factories and other places, the presence or absence of faults or failures is often based on sounds observed at specific locations, and a system that observes the sounds at relevant locations in a noisy environment and notifies of abnormalities is useful, so various methods have been proposed [1]. Scalograms, which are images of information obtained by wavelet transformation, are used to understand and extract feature values, but there are few examples of directly inputting this into machine learning, and there are no examples of this in a surround environment such as a factory.

We propose an audio compression method for machine learning in this paper. We have proposed a method in which audio data of inside a factory is processed to image data using wavelet transformation, its characteristics are clarified using some processes, and then machine learning is applied [2-4]. By grasping the characteristics from the time variation of the frequency characteristics, the accuracy of judging normality or abnormality could be improved. The important thing in this system is the size of the audio data to make it easier to transmit. A feature of the proposed method is the evaluation criteria for audio compression. The evaluation criteria of audio compression are based on the learning accuracy of machine learning, rather than the traditional judgment by human ears. Data compression is performed so as not to remove features that are important for machine learning in the proposed method.

Figure 1 shows the overall figure of the system we are developing. In the first stage, machine learning is performed using only normal sounds.

First, we apply wavelet transformation to each sound, visualize each sound data. Second, feature extraction processes are applied to the image data. This becomes the input data for machine learning. The wavelet transform uses a basis that shortens the time width at high frequencies and widens the time width at low frequencies, thus providing local frequency information and efficient time-frequency analysis. Time domain is essential for environmental sounds

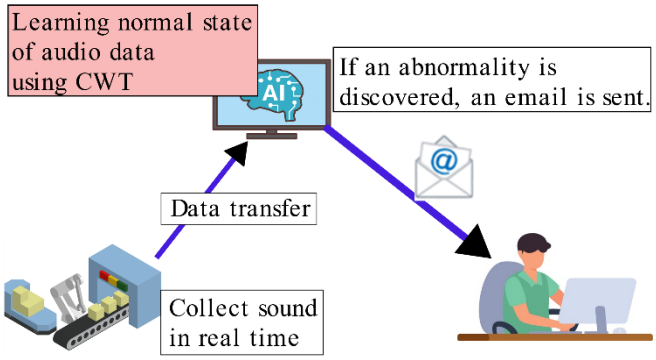


Figure 1: Proposed remote monitoring type system configuration for anomalous sound detection.

and the operating sounds of factory machinery. The AI learned from this input data is used to analyze the voice data collected in real time, and if an event occurs in which abnormal voice is observed, an alert is sent to the administrator (by email, etc.).

An important point in our system is how to collect and transfer audio data. There are multiple machines in a factory and multiple factories, but it is not realistic to install a server for each machine and each factory for the collected data. It is desirable to transfer the collected data in real time and consolidate it on one server. In such a situation, the size of the voice data is important. The smaller the size of the voice data, the easier it is to collect and transfer. However, this requires that the compressed voice data can be analyzed to determine whether it is a normal sound or an abnormal sound. This will make it possible in the future to build a system that stores collected data on the cloud, as shown in Fig. 2, and allows for real-time understanding of the situation from remote locations.

In this paper, we examined the compression method of audio data for anomalous sound detection system in factories for early failure detection using machine learning, thus we will report. In addition, discrimination was performed using images that applied wavelet transform as data for machine learning for abnormal sound detection.

When detecting anomalies using this method, the data is transferred and then analyzed to determine whether or not there is an abnormality; therefore, the proposed system is not suitable for capturing the moment an abnormality occurs and issuing an alarm. However, since the proposed system replaces the work of engineers, who listen to the sounds made by machines when they perform specific actions during inspections and judge the situation, it does not support immediate detection of abnormalities.

In this study, we hypothesize a system that observes and monitors sound, compresses and analyzes the sound using wavelet transform, and aims to verify the possibility and usefulness of applying this to sound source preservation and anomaly detection.

2 WAVELET TRANSFORM

There are two types of wavelet transform. One is continuous wavelet transform(abbreviated as CWT in the following text and figures) and the other is discrete wavelet transform(abbreviated as DWT in the following text and figures).

The wavelet transform is known as a method for analyzing both time and frequency. The wavelet transform is a method that calculates accurate frequency information by increasing the time width in the low-frequency region and calculates accurate time information by decreasing the time width in the high-frequency region. In the wavelet transform, the original waveform is expressed as an appropriate waveform $\Psi(t)$. This $\Psi(t)$ is called the mother wavelet, and the one that is appropriate for the waveform to be analyzed is selected appropriately. The mother wavelets used in this study are Morse, Molet, and Bump for the CWT, and SYM4 and DB4 for the DWT [5].

The CWT has applications such as detecting abnormal signals, and the DWT is used as a standard for image compression, and its applications are being actively discussed.

A typical analysis method used for audio signals is the Fourier transform, which includes frequency information but loses time domain. In contrast, the wavelet transform preserves time information, making it possible to analyze even sudden signal fluctuations. It is also possible to perform flexible operations such as varying the time interval according to the frequency domain, making it possible to perform dynamic analysis according to the situation.

Regarding the selection of the mother wavelet to be used in DWT, Ref. [6] shows the usefulness of Daubechies Wavelet (DB). Also, Ref. [7] shows the usefulness of Symmlet Wavelet (SYM). In this study, we confirmed the usefulness of actually recorded audio data, which may contain environmental sounds not mentioned in the previous studies, and showed the dependency of the compression rate on the parameters.

Waveforms indicating machine abnormalities often show signals with a sudden increase in waveform amplitude. This is similar to the waveform of an electrocardiogram. The SYM4 and DB4 wavelets are particularly suitable for analyzing biological signals, such as detecting the QRS complex in an electrocardiogram (ECG). The QRS complex is the most prominent feature of an electrocardiogram and reflects the deceleration of the left and right ventricles [8]. We therefore believe that SYM4 and DB4, which have been reported to be mother wavelets suitable for ECGs, will also be useful in this research.

In addition, we also verified the effectiveness of other mother wavelets that have been proposed and implemented in Matlab.

3 EXPERIMENT RESULTS: COMPRESSION OF AUDIO SIGNALS USING DISCRETE WAVELET TRANSFORM

3.1 Method of Compression Experiments

An experiment was conducted to confirm the effect of compression using DWT. The following items were confirmed:

1. Amount of file size reduction due to compression

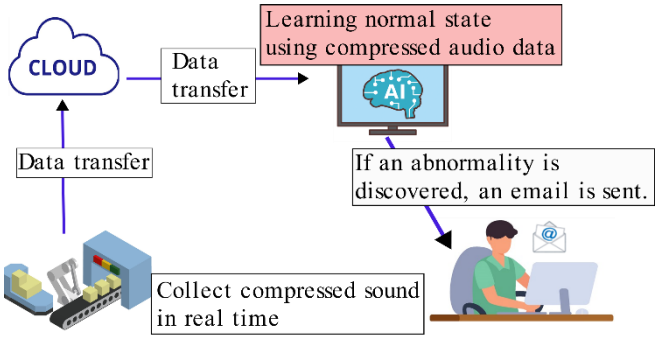


Figure 2: Proposed remote monitoring type system configuration for anomalous sound detection in the future.

2. Dependence of compression rate on mother wavelet
3. Dependence of compression rate on threshold used for analysis
4. Difference in compression rate due to difference in type of sound source

The audio data used in the experiment was collected from various sources [9-11], [12].

The original data was a 10-second uncompressed file with a sampling frequency of 16,000[Hz] . The sound sources include Fan, Pump, Slider, Toy Car, Toy Conveyor, Valve, and Toy Train, and each has a dataset of normal sounds and a dataset that includes abnormal sounds. For audio compression, we used the Wavelet Toolbox in MATLAB R2024.

3.2 File Size Reduction due to Compression

First, we analyzed the normal sound of a fan. We performed a DWT using SYM4 as the mother wavelet. Next, we applied a cut using the same threshold at each of the four frequency decomposition levels, and samples with values below the threshold were set to zero. Fig. 3 shows the chronological waveform diagram of the signal reconstructed based on the coefficient distribution of each decomposition level after cutting. The horizontal axis is the sampling number along the time series, and the vertical axis is the signal amplitude.

The reconstructed data was output as an audio file (wav file), and the audio file was converted to a flac file and compressed.

Figure 3 shows a comparison of the signal waveforms of a normal fan sound before (original) and after (compressed) DWT conversion. The horizontal axis is the number of samplings, which corresponds to time. The vertical axis is the signal amplitude. The first 1000 samples of the sound data are displayed, which corresponds to the first 62.5 milliseconds when converted to time.

The compression effect was evaluated by the compression ratio R. R is calculated as the ratio of the file size when the original audio data was converted to a flac file to the file size when it was converted to a flac file after DWT.

As we continued our analysis, we found that there was audio that had no effect at all from compression. Such audio has a waveform like that shown in Fig. 4. This waveform resembles a giant white noise.

In this case, the compression rate R was 1, meaning no compression was possible. Sounds like this are thought to be found frequently in sound sources such as fans, where the signal is expected to fluctuate at a high frequency. Since this type of waveform is thought to exist in both normal sound data and data containing abnormal sounds, we decided not to consider it noise this time, and to use the data as is for analysis.

The compression ratio R was calculated for 50 normal sound data and 50 abnormal sound data, and the average value was calculated. The error is the standard deviation calculated using the 50 values. For normal fan sounds, the compression rate was $R = 0.820 \pm 0.057$. When abnormal sounds were included, the compression rate was $R = 0.872 \pm 0.042$. SYM4 was used as the mother wavelet, and $T = 0.0017$ was used as the threshold for DWT compression.

This shows that audio data including noise and environmental sounds can be compressed using DWT.

3.3 Dependence on Different Types of Mother Wavelet

In the previous section, SYM4 was used as the mother wavelet for DWT, but we investigated the difference in compression ratio when using DB4, whose usefulness was shown in Ref. [6], [8], and other mother wavelets (bl7, beyl, coif1, fk4, haar, han5.5, mb4.2, vaid). In this case, we used the same threshold value T as in the previous analysis, T=0.0017. The results are shown in Fig. 5.

From this, we found that even with the same sound source, the compression rate varies greatly depending on the type of mother wavelet applied. For han5.5 and vaid, we found that although there was a large error, a low compression rate could be expected. Furthermore, for the Fan data set, SYM4 and DB4 were confirmed to be useful, as shown in Ref. [6] and [7]. It was also found that other mother wavelets can be expected to be useful.

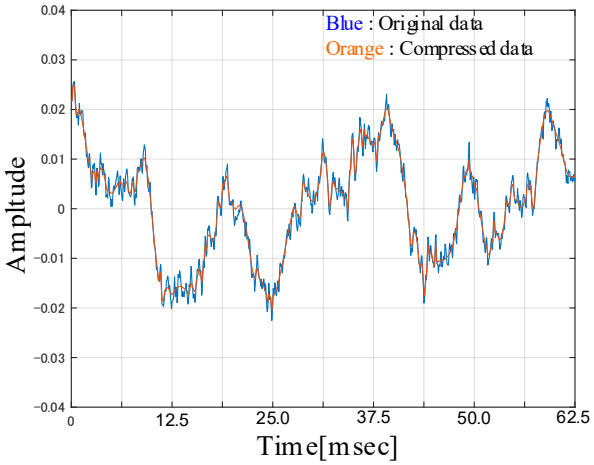


Figure 3: Comparison of Waveform diagram of Normal data using SYM4-Wavelet.

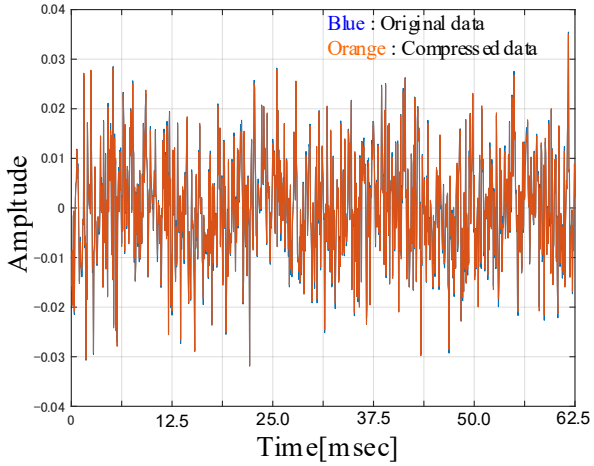


Figure 4: Comparison of Waveform diagram of Normal data using SYM4-Wavelet.

3.4 Dependence of Compression Rate on Threshold used for Analysis

The threshold value (T=0.0017) for compression with DWT was determined by performing compression using Matlab's Wavelet Signal Analyzer. We investigated the dependency of this value on the compression ratio R.

Using the Fan dataset as the sound source, Fig. 6 shows the change in compression ratio R when the threshold is changed. 50 pieces of data were analyzed for each and the average value was calculated.

Looking at this, we can see that the default value (T=0.0017) is exactly at the midpoint where the data size changes rapidly. Increasing the threshold decreases the compression rate R, but when we checked the played back waveform, we could see that it became increasingly distorted.

When compressing with DWT, the determination of the threshold is important. In the analysis that follows, we plotted the compression ratio R vs. the threshold, as shown in the figure, and derived a polynomial fit curve for the graph. Using this equation, we determined the threshold when R was 0.8, and used this value for the analysis. If we were actually archiving audio from a particular factory system, we would need to accurately measure this dependency on the threshold and optimize the value using the methodology used here.

3.5 Difference in Compression Rate due to Difference in Type of Sound Source

To confirm the contents of the previous section, we investigated how the compression rate differs depending on the type of sound source(Fan, Pump, Slider, Toy Car, Toy Conveyor, Valve, and toytrain) when the same threshold is set.

Figure 7 shows the compression rate of each sound source when the threshold is set to T=0.0025. 50 pieces of data for each were analyzed and the average value was calculated.

From this, we found that toytrain and other sound data have an extremely high compression effect.

4 EXPERIMENTAL RESULTS: CONTINUOUS WAVELET TRANSFORM OF COMPRESSED AUDIO DATA

4.1 Comparison of Raw Data and Compressed Data

Next, a CWT was performed on both the normal and abnormal sounds in the compressed voice data to create a scalogram that represents the signal strength on the time and frequency plane.

Figure 8 shows the scalogram of Normal raw data where Morse was used as Mother Wavelet. Fig. 9 shows the scalogram on Compressed data using the DWT of SYM4 Wavelet.

From this, it was found that by first compressing using the discrete wavelet transform, the components corresponding to the white noise that are common to both were suppressed, resulting in a distribution that makes it easier to determine the characteristics.

The abnormal data were analyzed in a similar manner. The results were shows in Fig. 10 and 11. Abnormal data has a high retention rate, but it is observed as a more characteristic figure.

4.2 Dependence on Motherwavelet

To investigate the dependence on Motherwavelet, a similar analysis was performed for two types of Motherwavelet, amor(Morlet) and Bump, in addition to the basic Morse.

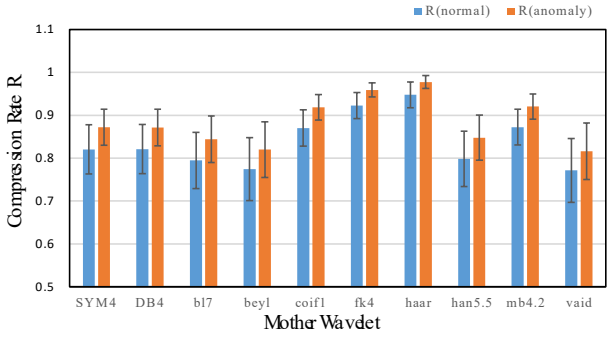


Figure 5: Comparison of Compression Rate R for different mother wavelets.

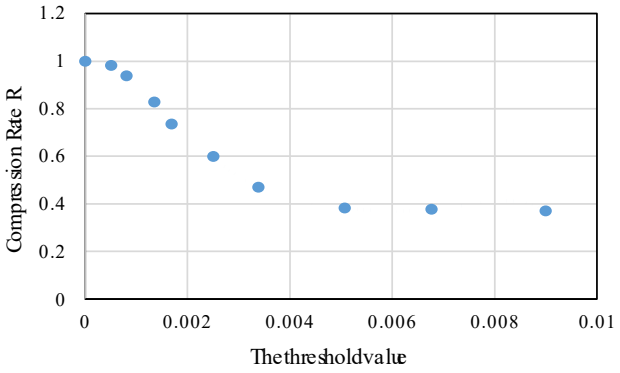


Figure 6: Comparison of Compression Rate R for the threshold value for compression with DWT.

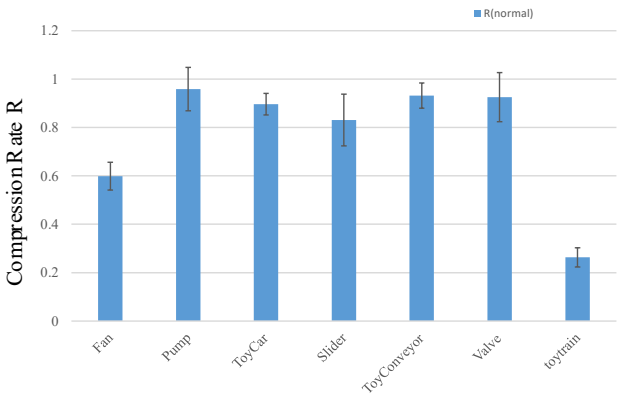


Figure 7: Comparison of Compression Rate R for depending on the type of sound source.

Figure 9 shows the case where the basic Morse-Wavelet were used as Motherwavelets for Normal data analysis, Fig. 12 shows amor(Morlet) and Fig. 13 shows Bump-Wavelet.

From Fig. 12, this scalogram was more widely distributed than in the case of Morse and Bump. So, Morlet was not suitable to handle this data.

From this, it was found that for continuous wavelets, differences in the scalogram appear depending on the mother wavelet.

The abnormal data were analyzed in a similar manner. The results were shows in Fig. 11 ,14 and 15. For Abnormal data, the distribution characteristics were completely different. By comparing Fig. 11 and Fig. 14, as expected, Morlet(amor) was found to be inappropriate for extracting features from this data.

On the other hand, in Fig. 15, the scalogram using Bump-Mother Wavelet, it was found only in a few places, and was very distinctive. Fig. 16 shows the scalogram of uncompressed data transformed by CWT using Bump. We can see that Bump is a wavelet that is quite suitable for extracting features from this Abnormal data.

This can also be derived by comparing Fig. 14(Morlet-Wavelet) and Fig. 15(Bump-Wavelet).

So, it is useful to first identify the wavelet that most clearly shows the characteristics of the target sound source.

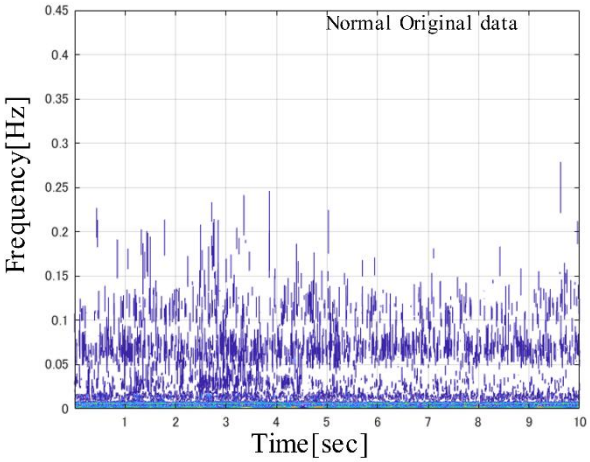


Figure 8: Scalogram of Normal Original Data.

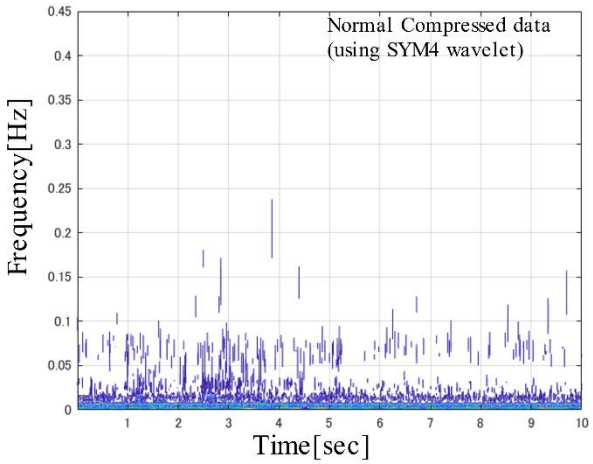


Figure 9: Scalogram of Compressed Normal Data(compressed using SYM4 Wavelet).

Since the range of the distribution of the scalogram is the very characteristics of the sound source, it is possible to use these distribution maps as data for machine learning to build a system that detects and judges abnormalities on behalf of humans.

These results confirmed that the proposed method makes it

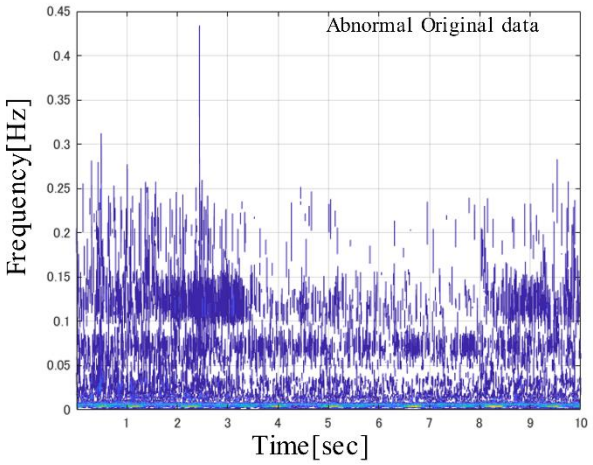


Figure 10: Scalogram of Abnormal Data.

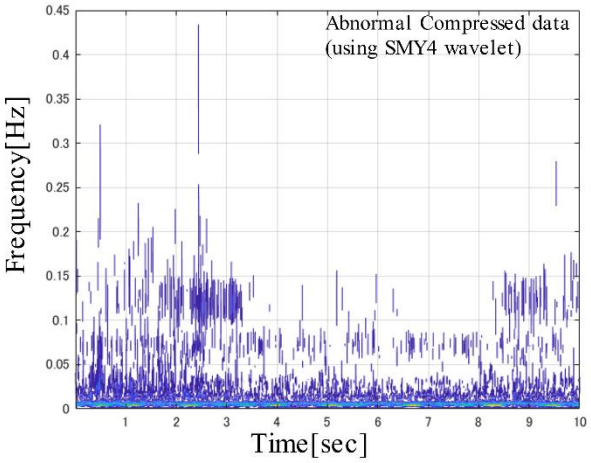


Figure 11: Scalogram of Compressed Abnormal Data(compressed using SYM4 Wavelet).

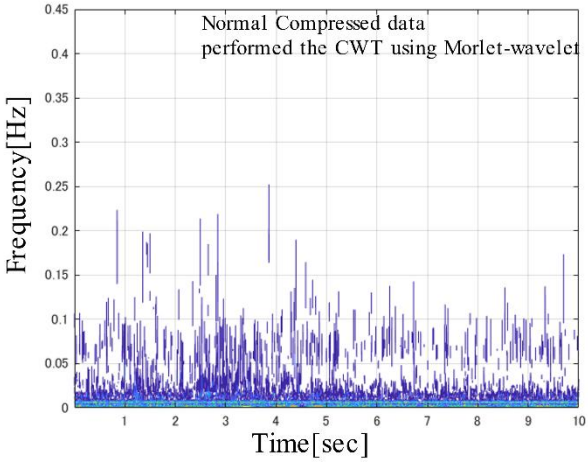


Figure 12: Scalogram of Compressed Normal Data for Morlet-Wavelet (compressed using SYM4 Wavelet).

possible to visually determine whether data is normal or abnormal even after compression, and that the file size can be reduced to 70% of original data.

5 DISTINGUISHING BETWEEN NORMAL AND ABNORMAL USING MACHINE LEARNING WITH SCALOGRAM IMAGES

5.1 Discriminating Audio Signals in a Surround Environment

We used MATLAB R2024a to perform machine learning using scalogram images as input data.

We used VGG16, a neural network with pre-trained weights for image recognition, to perform transfer learning on scalograms. As a dataset, we used 300 images each of normal and abnormal sounds. The trained network was used to classify audio data, and the percentage of data that was successfully classified was used as the validation accuracy. This calculation was performed three times with different random numbers, and the average value was calculated. As a condition for creating the scalogram, the mother wavelet used in the CWT was BUMP, which was found to be useful in the above analysis.

First, we calculated the validation accuracy for data that did not undergo DWT. However, since we assume that machine learning judgments will be made on audio data after it has been compressed using DWT, we also performed machine learning on the compressed data to calculate the validation accuracy.

SYM4 was used as the mother wavelet for DWT.

The types of audio data used were FAN [9-11], PUMP [9-11], and ToyTrain [12].

The results are shown in Table 1.

The validation accuracy for Fan and pump was very low. As previously reported, FAN contains a mixture of waveforms with a compression rate of 1, and various types of sounds are mixed into what is considered normal sound. Therefore, when performing DWT, it is expected that verification accuracy will improve by optimizing parameters, particularly by changing the frequency resolution.

This analysis showed that it is possible to distinguish between normal and abnormal sounds even using data compressed with DWT.

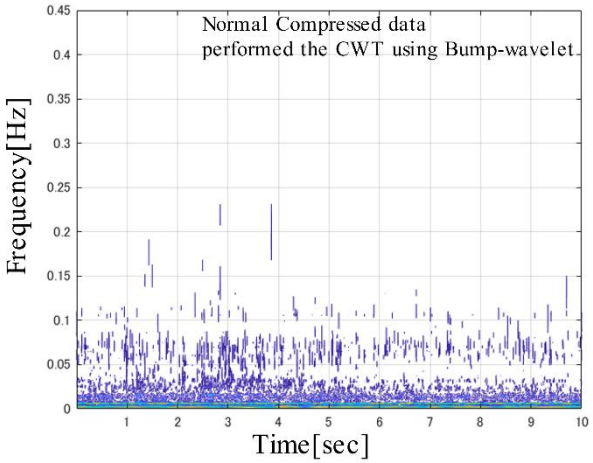


Figure 13: Scalogram of Compressed Normal Data for Bump-Mother Wavelet(compressed using SYM4 Wavelet).

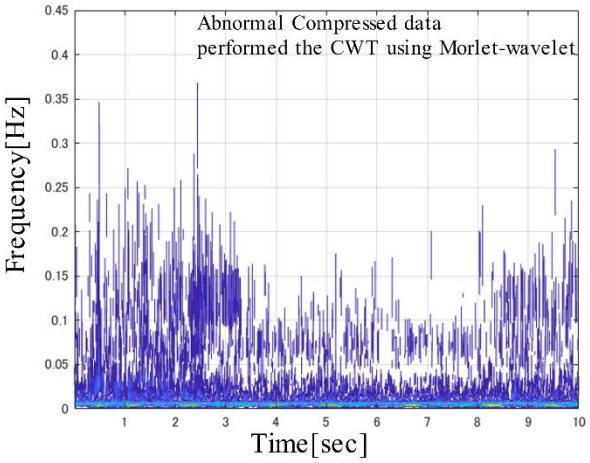


Figure 14: Scalogram of Compressed Abnormal Data for Morlet-Wavelet(compressed using SYM4 Wavelet).

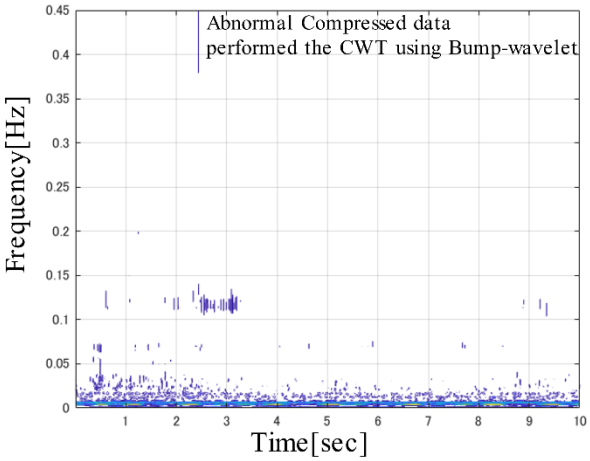


Figure 15: Scalogram of Compressed Abnormal Data for Bump-Wavelet(compressed using SYM4 Wavelet).

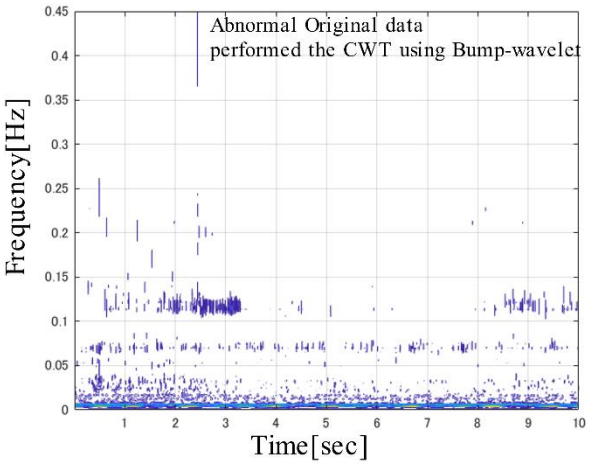


Figure 16: Scalogram of Abnormal Original Data for Bump-Wavelet(Uncompressed data).

It was also found that compression using DWT improves validation accuracy.

5.2 Comparison with Existing Audio Compression Methods

Next, we investigated whether there was a difference in the validation accuracy of machine learning depending on the compression method used to create the scalogram image to be input into machine learning.

To verify subtle differences in validation accuracy, we used the toytrain data, which had high validation accuracy in the previous analysis, changed the random numbers, calculated the validation accuracy 15 or 20 times, and calculated the average value.

Scalograms were created using uncompressed audio signals, audio signals converted to mp3 format (8 kbit/s), audio signals converted to mp3 format (16 kbit/s), and audio signals compressed using DWT (using SYM4-wavelet). Machine learning was then performed using these images as input to determine the verification accuracy for distinguishing between normal and abnormal audio.

The results are shown in Table 2.

This shows that the validation accuracy of anomaly detection using data compressed by DWT is higher than when data is compressed using other methods. Furthermore, it is possible to obtain values equivalent to those obtained when detection is performed using uncompressed data.

Figure 17 is a scalogram of the fan's voice, while Figure 18 is a scalogram of the data compressed with MP3. Comparing the two, it can be seen that Fig. 18 has mostly lost its high frequency range. As a result, information corresponding to the features in the image disappears, which is thought to reduce verification accuracy.

These analyses show that the use of wavelet transform is a useful compression method that can maintain the learning accuracy of machine learning.

6 SUMMARY

In this study, data compression was performed using discrete wavelet transform with SYM4, DB4 and some other mother wavelets. As a result, we found that using wavelet transform is useful as a compression method that can maintain the learning accuracy of machine learning. In addition, by performing continuous wavelet transform on compressed data, we were able to obtain characteristic scalograms in some cases.

We found that it is possible to treat the scalogram of a compressed audio signal as a simple two-dimensional image and use it as input data for machine learning to distinguish between different situations.

We also found that this method can maintain validation accuracy in machine learning.

Table 1: Comparison of validation accuracy using machine learning.

	validation accuracy(%)	
Audio Data	Without Compression	Compressed data
FAN	61.8	67.4
PUMP	89.9	91.1
ToyTrain	99.6	99.7

Table 2: Comparison of validation accuracy using different compression methods.

	Without Compression	Compressed data		
		MP3 8kbit/s	MP3 16kbit/s	DWT
validation accuracy (%)	99.6	95.9	97.7	99.6

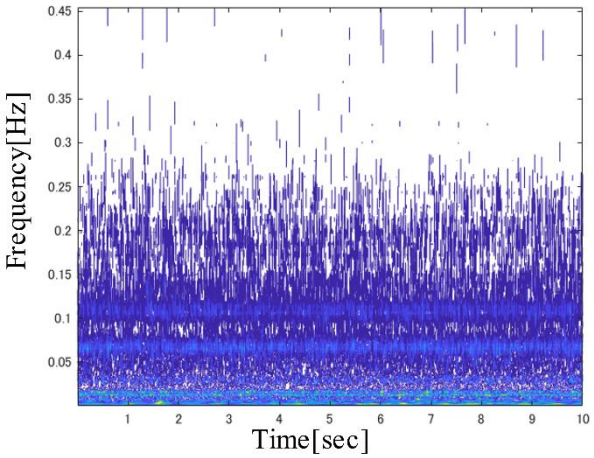


Figure 17: Scalogram of Normal Fan Data without compression.

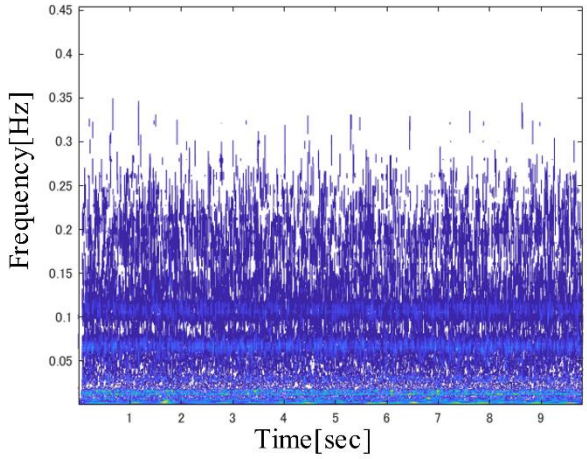


Figure 18: Scalogram of Normal Fan Data with compression using MP3.

ACKNOWLEDGMENT

This research is supported by Sumitomo Metal Mining Co., Ltd.

REFERENCES

[1] E.Awada, A.Al-Qaisi, E.Radwan, M.Nour, “Motor fault detection using sound signature and wavelet transform,” International Journal of Power Electronics and Drive Systems (IJPEDS),vol.13,pp247-255(2022).

[2] M.Ueda, D.Tanaka, M.Sode, “A study on anomalous sound detection in factories for early failure detection using wavelet transform,” IEEE International Conference on Consumer Electronics – Taiwan, 2024 (IEEE ICCE-TW),July 9-11, 2024(2024).

[3] K.Kikuchi, M.Sode, “Anomalous Sound Analysis Using Wavelet Transform,” The 7th NIT-NUU Bilateral Academic Conference 2024, July 14-15, 2024(2024).

[4] H.Yamamoto, M.Sode, “Anomaly Analysis Using Wavelet Transform,” The 7th NIT-NUU Bilateral Academic Conference 2024, July 14-15, 2024(2024).

[5] Choose a Wavelet, https://jp.mathworks.com/help/wavelet/gs/choose-a-wavelet.html(last visited: 07.20.2024).

[6] M.S.Serina, S.G.Mosin,”Digital Audio Information Compression Using Wavelets”,IEEE Xplore https://doi.org/10.1109/MIXDES.2006.1706666, “MIXDES 2006 Gdynia,POLAND 22-24 June 2006”(2006).

[7] R.C.Guido, C.D.Maciel, M.Monteiro, E.S.Fonseca, S.Panchapagesan, J.C.Pereira, L.S.Vieira, S.Barbon Jr, M.A.B.Guilherme, K.I.C.Sergio, T.L.Scarpa, P.C.Fantinato, E.J.Rodrigues de Moura,” A Study on the Best Wavelet for Audio Compression”, IEEE Xplore https://doi.org/10.1109/ACSSC.2006.355141,”ACSSC 2006 ”(2006).

[8] Wavelet analysis of biological signals, https://jp.mathworks.com/help/wavelet/ug/wavelet-analysis-of-physiologic-signals.html(last visited: 07.23.2024).

[9] K.Dohi, K.Imoto, N.Harada, D.Niizumi, Y.Koizumi, T. Nishida, H.Purohit, T.Endo, M.Yamamoto, Y.Kawaguchi, Description and Discussion on DCASE 2022 Challenge Task 2: Unsupervised Anomalous Sound Detection for Machine Condition Monitoring Applying Domain Generalization Techniques. In arXiv e-prints: 2206.05876(2022).

[10] K.Dohi, T.Nishida, H.Purohit, R.Tanabe, T.Endo, M. Yamamoto, Y.Nikaido, and Y. Kawaguchi. MIMII DG: sound dataset for malfunctioning industrial machine investigation and inspection for domain generalization task. In arXiv e-prints: 2205.13879(2022).

[11] N.Harada, D.Niizumi, D.Takeuchi, Y.Ohishi, M.Yasuda, and S.Saito. ToyADMOS2: another dataset of miniature-machine operating sounds for anomalous sound detection under domain shift conditions. In Proceedings of the 6th Detection and Classification of Acoustic Scenes and Events 2021 Workshop (DCASE2021), pp.1–5. Barcelona, Spain, November 2021(2021).

[12] Y.Koizumi, S.Saito, N.Harada, H.Uematsu and K.Imoto, "ToyADMOS: A Dataset of Miniature-Machine Operating Sounds for Anomalous Sound Detection," in Proc of Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), 2019(2019).

(Received: November 15, 2024)
(Accepted: August 13, 2025)

Miyuki Shirai.
She received the Ph.D in Science from the Graduate School of Science and Technology, Niigata University, and has been engaged in research on high-energy physics at the National Institute of Technology (KOSEN), Niihama College. She is currently working on simulations related to radiation and radiation sensors. She is interested in applying AI for signal analysis.

Yuhi Shuno
In 2025, he graduated from the Department of Electrical and Information Engineering at Niihama College of Technology.In the same year, he enrolled in the Advanced Course in Electronic Engineering at the same institution.His research focuses on anomalous sound detection using audio and vibration signal processing and machine learning, with an emphasis on wavelet-based feature extraction for rotating machinery diagnostics.

Hiroki Yamamoto
He received his Associate Degree in Engineering from the Department of Electrical and Computer Engineering, National Institute of Technology (KOSEN), Niihama College, Japan, in 2025. He joined Central Japan Railway Company (JR Central) in the same year.

Sho Ishikawa
He received his Associate Degree in Engineering from the Department of Electrical and Computer Engineering, National Institute of Technology (KOSEN), Niihama College, Japan, in 2025. He is currently pursuing a Bachelor’s degree at the School of Computer Science and Engineering, Toyohashi University of Technology.

Mikiko Sode
She was engaged in the development of supercomputers, including ACOS systems and the Earth Simulator, at NEC Corporation. After moving to Renesas Electronics Corporation, she worked on the development of automotive LSIs such as the R-Car series. She is currently involved in research on transforming everyday infrastructure, such as bus stops, into IoT-enabled systems and utilizing edge computing technologies to support life-oriented services including community monitoring. She received the Ph.D. degree in Engineering from Waseda University. She is a member of the Information Processing Society of Japan (IPSJ), IEEE, and the Institute of Electronics, Information and Communication Engineers (IEICE).

Practical Paper

A Proposal of Smartphone Beacons in Stay Estimation System Using BLE

Kota Togawa†, and Katsuhiko Kaji†

†Graduate School of Business Administration and Computer Science, Aichi Institute of Technology
{b24716bb, kaji}@aitech.ac.jp

Abstract - People spend 88.9% of their day indoors and are mainly active indoors in physically separated spaces such as their own rooms, laboratories, and conference rooms. Therefore, room-level location information, rather than highly accurate location information, is also valuable. We have proposed a stay estimation system that receives signals from BLE beacons carried by each user and estimates the room location using receivers installed in the environment. However, conventional methods using only physical beacons have problems such as battery replacement, time-consuming initial setup, and users moving from room to room without a physical beacon. In this study, we implement and evaluate smartphone beacons that have high tracking performance, do not require initial setup or battery replacement, and consumes little battery power. Smartphones are often carried around at all times, so they are considered to be highly trackable. We implemented an application that automatically sets the content of physical beacon advertisements and continues to advertise BLE signals periodically. In evaluation experiments, the application was found to have high tracking performance and low battery consumption.

Keywords: Room-level indoor localization, BLE beacon, Smartphone application

1 INTRODUCTION

Since many indoor activities take place within the confines of rooms, room-level indoor location information is valuable. People spend the majority of their time indoors, as reported in the study that people spend 88.9% of their day indoors [1]. Indoors, people are mainly active in physically separated spaces such as their own rooms, laboratories, and conference rooms. Therefore, not only highly accurate location information but also room-level location information is valuable. For example, there are commercial packages available that use room-level location information to manage attendance [2], monitor congestion, and reserve meeting rooms [3].

A method using radio signals is available for room-level location estimation (hereafter referred to as "room-level localization"). Among such methods, those using BLE signals use dedicated devices that transmit BLE signals (hereinafter referred to as "physical beacons") and a receiver that receives BLE signals. The receiver receives BLE advertising packets from nearby physical beacons. The BLE advertising packets contain the MAC address and UUID. This information can determine which physical beacon the receiver received the BLE signal from. Room-level localization using BLE signals is based on this property.

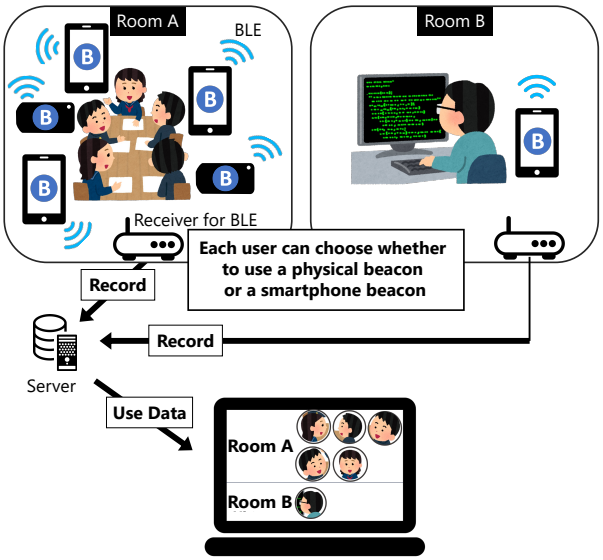


Figure 1: Schematic diagram of the stay estimation system to be realized in this study. The user can choose whether to use a smartphone beacon or a physical beacon.

There are two main types of room-level localization methods using BLE signals. The first one is a method in which BLE signals from physical beacons installed in rooms are received by a device such as a smartphone carried by each user and sent to a server. While this method does not require each user to carry exclusive devices, it does require all users to install an application on their own devices. In addition, there is a concern that battery consumption will increase because the application is always running. The second one is to receive BLE signals from the physical beacon carried by each user with a receiver installed in rooms and transmit them to the server. In this method, the users only need to carry the physical beacon distributed by the administrator. We have proposed the system called "StayWatch" that manages the stay information of the members of our laboratory, and we are operating it in our laboratory. The purpose of StayWatch is to activate communication among multiple communities by sharing privacy-conscious stay information.

There are also several issues with the method where the users carry a physical beacon. Since each user carries one physical beacon, the number of beacons needed to be distributed must match the number of users in the community

	Smartphone Beacon	Physical Beacon
User Burden	Install the application on the user's own smartphone.	Battery needs replacement (About once a year)
Administrator Burden	None	Initial setting of UUID, etc.
Financial Cost	Low (Use user's smartphone)	High (Thousands of yen per beacon)
User Restrictions	No restrictions on app installation	Available to everyone

Figure 2: Features of smartphone beacons and physical beacons

where the beacons are to be deployed. The financial cost is high because many physical beacons are required. In addition, it takes a lot of time and effort to set up physical beacons. When performing room-level localization, individuals are identified using UUIDs advertised from physical beacons. When installing in a community, it is necessary to manually enter and set the UUID of the physical beacon for each user. Other problems include the fact that physical beacons are sometimes placed in rooms and moved from room to room, resulting in low tracking performance, and the fact that some users faling to notice when the batteries in the physical beacons run out, or neglecting to replace the batteries even if they do notice.

The purpose of this research is to realize a BLE beacon with high tracking performance and low financial and human costs. As an approach, we implement and evaluate an application that substitutes the role of a physical beacon with a smartphone (hereinafter referred to as "smartphone beacon"). Smartphone beacons are expected to offer better tracking performance than physical beacons, as many people carry their smartphones with them at all times. An overview of this application is shown in Fig. 1. While smartphone beacons have multiple advantages as shown in Fig. 2, they also have disadvantages such as the need for each user to install the application on their own device. While physical beacons have multiple disadvantages, they also have the advantage that users only need to carry a physical beacon distributed by the administrator, which is less burdensome. In this way, there are advantages and disadvantages for both smartphone beacons and physical beacons. Therefore, the system should be configured so that each user can choose whether to use a smartphone beacon or a physical beacon.

This study makes two contributions. The first contribution is the proposal of a system configuration that enables smartphones and physical beacons to coexist within the same presence estimation system, and the demonstration of its practical feasibility through evaluation experiments. Separate systems existed where users carried either smartphones or physical beacons. However, there was no configuration allowing users to choose whether to carry a smartphone or a physical beacon. The second contribution is the clarification of the advantages of using smartphones as transmitters compared to using them as receivers. Using smartphones as receivers reduces financial costs since no physical beacons need to be prepared for each user, and it also makes third-party tracking more difficult. Therefore, the receiver-based approach has been more widely adopted than the transmitter-based approach. Within this context, this study highlights the advantages of using smartphones as transmitters. Smartphones consume less battery power when transmitting beacon signals than when receiving them. Moreover, because they can operate with the same protocol as physical beacons, coexistence is easier to achieve.

2 RELATED RESEARCH

There are various methods for room-level localization using Wi-Fi, BLE, IC cards, cameras, voice recognition, etc. Room-level localization can also be achieved using only information on entry and exit from a room. Methods that take advantage of this are those using IC cards, cameras, and voice recognition. In the method using IC cards, readers are installed at entrances and exits, and users hold their student ID cards or other IC cards over the readers when entering or exiting a room [4]. This requires the user to take active action, which causes operation forgetting. There are methods that do not require active action, such as the voice recognition method where the user says his or her name when entering or exiting a room [5]. Another commercial package is available that uses facial recognition by cameras installed at entrances and exits [6]. However, these methods face issues with reduced accuracy due to users with similar names or poses, as well as the angle of the face captured by the camera.

There are room-level localization methods that use real-time radio signal information, such as Wi-Fi and BLE, as well as entry/exit determination. Some Wi-Fi-based methods detect and use packets sent by smartphones to locate Wi-Fi base stations in the surrounding area [7]. Users do not need to install special applications on their device. However, in recent years, MAC addresses of smartphones are often randomized to improve privacy, in which case this method cannot be used. The fingerprinting method is a method that enables highly accurate location estimation using radio waves [8]. The fingerprinting method has high administrative costs. Room-level localization does not require as much accuracy as the fingerprint method. Therefore, the proximity method, which estimates location at the area level, is often used for room-level localization.

BLE allows for easier adjustment of the location and number of base stations than Wi-Fi. Wi-Fi base stations are installed to improve the wireless communication environment. Therefore, it is difficult to relocate or increase the number of Wi-Fi base stations solely for the purpose of improving the accuracy of room-level localization. In contrast, physical beacons that transmit BLE signals are installed specifically for room-level localization, making it easier to place them in optimal locations and to adjust the number of beacons.

There are two main methods for room-level localization using proximity to BLE beacons: one is to install a physical beacon in rooms and the other is to have users carry physical beacons. In the method that installs physical beacons in rooms, physical beacons are first installed in each room. Each physical beacon should advertise a unique ID. Users run the application on their smartphones to receive signals from the physical beacons. When a user enters a room, their smartphone receives the signals from the physical beacons. The

unique ID contained in the signals from the physical beacon is sent to the server to determine which smartphone is being used and which room the physical beacon is in. In this way, the system can estimate who is in each room. Physical beacons do not require any cables; they can be simply fixed to a wall or ceiling. Users do not need to carry any special equipment. A disadvantage is that it is necessary to install a specific application on the smartphone. In addition, the smartphone must continuously receive BLE signals from the surrounding area and communicate with the server, which places a heavy burden on the smartphone and consumes a large amount of battery power.

In the method where each user carries a physical beacon [9], receivers are first installed in each room. The administrator distributes small physical beacons to users, which are configured to advertise a unique ID. When a user enters a room, the receiver in that room receives signals from the physical beacon the user is carrying. The unique ID contained in the signals from the physical beacon and the unique ID of the receiver are sent to the server. The server checks whose physical beacon it is and which room the receiver is in. In this way, it can estimate who is in which room. Users only need to carry the distributed physical beacon in their wallets, bags or similar items. The disadvantage is that the financial cost is high because physical beacons are required for each user. In addition, the initial configuration of physical beacons is time-consuming because it is done manually for each beacon using a configuration application.

When multiple rooms are assumed, depending on the material and thickness of the walls, BLE signals may penetrate walls. In such cases, comparing the signal strength from receivers or BLE beacons installed in each room enables more accurate room estimation.

There is a study that uses BLE signals emitted from smartphones to estimate the level of congestion in a room [10]. For this purpose, the COVID-19 contact tracing application (hereafter referred to as COCOA) [11], which was provided by the Ministry of Health, Labour and Welfare, is used. Smartphones with COCOA installed transmit BLE signals periodically. This BLE signal is received by a receiver installed in the room. The number of smartphones in the surrounding area is determined based on the information about how many BLE signals the receiver receives from them. In this way, the number of people in the room is estimated. Users only need to install COCOA on their smartphones and do not need to install any additional applications for the stay estimation system. Users who do not wish to install COCOA can be provided with physical beacons to carry instead. However, COCOA cannot identify individuals because the contents of advertised packets are encrypted and updated periodically. Therefore, COCOA cannot be incorporated into a stay estimation system.

The acquired stay information can be used in various situations, contributing to the realization of a smart city [12]. For example, it can be used to estimate the room occupancy rates for indoor disaster rescue activities and power management. Another study has been conducted that promotes laboratory visits by predicting and presenting future stay information based on each laboratory member's past stay data and calendar schedules [13].

3 SMARTPHONE BEACONS IN STAY ESTIMATION SYSTEMS

In this chapter, we discuss the requirements specification and implementation of the smartphone beacon proposed in this study. Users choose whether to use a physical beacon or a smartphone beacon and carry the chosen device. Receivers in the room receive BLE signals transmitted from the beacon device and send them to the server to record stay information.

3.1 Examination of Requirement Specifications

Based on the background of this research presented in Chapter 1 and related research presented in Chapter 2, the requirements for a smart beacon are as follows.

(R1) Each user can choose whether to use a smartphone beacon or a physical beacon

(R2) Small battery consumption of the smartphone

(R3) Small initial setup effort

(R4) Users do not need to open the application except when changing settings

R1 : In order to allow the user to choose whether to use the smartphone beacon or the physical beacon, the smart phone beacon is made to behave in the same way as the physical beacon. It is appropriate for users to be able to choose between smartphone beacons and physical beacons, as each has its own advantages and disadvantages. Physical beacons periodically advertise packets containing a UUID to identify the user. Smartphone beacons should behave in the same way. This would eliminate the need for major modifications to the existing stay estimation system.

R2 : In order to suppress battery consumption, the application only transmits BLE signals during normal operation. It is necessary to make it a low battery consuming application, because smartphone beacon operates constantly. Periodic transmission of BLE signals can continue for several months to several years with a coin battery. Thus, transmitting BLE signals consumes very little power. Therefore, as in the solution of **R1**, battery consumption can be suppressed by having the smartphone beacon operate in the same way as the physical beacon. The network communication with a smartphone consumes a large amount of battery power. Therefore, the smartphone beacon does not require network communication except when it is set up, and its normal operation is limited to transmitting BLE signals, which consume low battery power.

R3 : To reduce the time and effort required for initial setup, the application automatically performs the setup. A community administrator performs the initial configuration of the physical beacon using a configuration application and manually enters the UUID and transmission frequency. The initial setup takes more than one minute per beacon, and about one in every 20 beacons is incorrectly configured. Additionally,

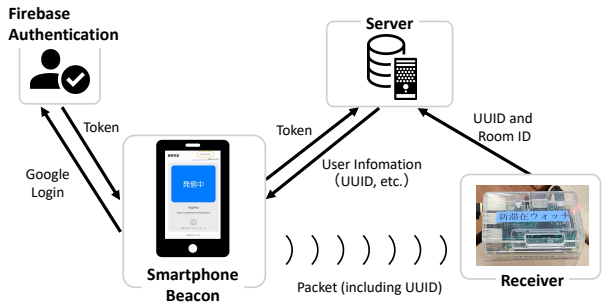


Figure 3: Initial setup flow

while the initial setup of physical beacons is done by the administrator, it is not desirable for the administrator to directly operate and configure the smartphone beacon, which is installed on each user's smartphone. We believe that even if the user performs the initial setup, it takes a long time for a person who is familiar with the initial setup to make a mistake, and that an unfamiliar user will take more time and make a mistake. The smartphone beacon automatically performs the initial setup, to reduce this burden on users.

R4 : To prevent users from forgetting to operate the application, they should not need to open it except during initial setup. Applications that need to be opened periodically require active user action. In the case of such an application, the application may not run due to forgotten operations or troublesome operations. Therefore, the user does not need to actively open the smartphone beacon except for the initial setup or when changing settings. Therefore, smartphone beacons should continue operating unless users intentionally stop it.

3.2 How to Set Up Outgoing Packets

Figure 3 shows the flow of automated initial setup to satisfy **R3**. The beacon uses OAuth authentication with a Google account to get the user's UUID. OAuth authentication prevents unauthorized acquisition of the user's UUID. At the first startup, the user logs in with their Google account according to the application screen. A token is issued upon successful authentication with Firebase Authentication. The smartphone beacon sends the token to the server that manages the user information. The server uses the information in the token to verify whether the user has authenticated with Firebase Authentication and whether the user is registered. If the server confirms that the token has been authenticated and the user has been registered, it sends the user information associated with the e-mail address to the smartphone beacon. The beacons use the user information got from the server to perform the initial setup and advertise a packet containing the UUID. In this way, the smartphone beacon satisfies **R3**.

3.3 Implementation in Android

We implemented the Android version of the smartphone beacon to satisfy **R1** and **R4**.

3.3.1 How to Transmit BLE Signals

To satisfy **R1**, smartphone beacons for Android advertise the UUID got from the server as it is included in the packet. Smartphone beacons transmit BLE signals using android.bluetooth.le package provided by Android. The smartphone beacons use this to set the UUID to be advertised, the frequency of transmission, and the transmission strength, and transmit BLE signals. However, the transmission frequency and strength cannot be strictly set, with only three levels of transmission frequency and four levels of transmission strength available. When the application is not open (hereafter referred to as "background"), the device transmits at the lowest frequency (ADVERTISE_MODE_LOW_POWER) for any of the settings. In order to accommodate large rooms, the transmission strength is set to the strongest setting ADVERTISE_TX_POWER_HIGH. In this way, the smartphone beacon satisfies **R1**.

To satisfy **R4**, the smartphone beacon must also run in the background. For this purpose, the smartphone beacons use a foreground service that allows applications to run in the background. This allows BLE signals to continue being transmitted even if the beacon is task-killed or the screen is put to sleep. However, it is not possible to satisfy **R4**, because this alone may stop the transmission of BLE signals.

3.3.2 How to Prevent BLE Radio Transmissions from Stopping

To satisfy **R4**, when the smartphone is unable to continue transmitting BLE signals, it will automatically resume when it is ready to transmit. There are three scenarios where BLE signals transmission in Android becomes uncontrollable: when the Bluetooth function is disabled, when the smartphone is turned off and the foreground service is terminated, or when the foreground service is forcibly terminated by Android.

When the smartphone becomes unable to use the Bluetooth function, the smartphone automatically resumes transmitting BLE signals when it becomes able to do so. When the user turns on Airplane mode, the Bluetooth function becomes unusable and BLE signals transmission stops. When Airplane mode is turned off, the smartphone beacon should automatically resume BLE signals wave transmission. For this purpose, we use the BluetoothAdapter.ACTION_STATE_CHANGED intent broadcasted by Android when the Bluetooth status of the smartphone is changed. When the foreground service detects this intent, it executes a process to start transmitting BLE signals.

When the smartphone is turned off and the foreground service is stopped, BLE signal transmission automatically resumes when the smartphone is turned on. When the smartphone is turned off, the foreground service stops and BLE signal transmission stops. When the smartphone starts up, the foreground service should automatically start and BLE signal transmission should resume. To achieve this, use the ACTION_LOCKED_BOOT_COMPLETED intent broadcast by Android when the smartphone is turned on. When the beacon detects this intent, it starts the foreground service and begins transmitting BLE signals.

When Android forces the application to stop, the BLE transmission of the smartphone beacon is automatically resumed when the resource is reused. If the foreground service has been running for a long time, its priority will gradually decrease, and it may be forced to terminate. In such cases, it is better for the beacon to automatically restart the foreground service and resume BLE signal transmission. To do this, set the return value of the method that is executed when the foreground service is started to START STICKY. When Android kills the foreground service, it will be automatically restarted as soon as the resources become available again. In this way, the smartphone beacon satisfies **R4**.

The smartphone beacon targets devices running Android 7.0 (API level 24) and later. We confirmed its correct operation on devices running Android 11.0(API level 30) through Android 15.0(API level 35).

3.4 Implementation in iOS

We implemented the iOS version of the smartphone beacon to satisfy **R1** and **R4**.

3.4.1 Restrictions on BLE Signal Transmission in The Background

iOS does not allow a user's UUID to be included in BLE advertisement packets as-is when the application is running in the background. When the application is in the foreground, it is possible to transmit using the iBeacon format, allowing the service UUID to be directly set to the user's UUID, just like on Android or with physical beacons. However, in the background, iOS alters the advertisement data. The iBeacon UUID is placed into an overflow area, resulting in a transformed value being included in the packet.

The value that changes according to the iBeacon UUID in the packet actually advertised by the smartphone beacon is a 152-bit hexadecimal number. Of the 152 bits, 24 bits are fixed values and 128 bits are values that change according to the set UUID. The initial value of the value that changes according to the iBeacon UUID is 0 for all 128 bits. Due to this transformation, the receiver cannot directly retrieve the iBeacon UUID from the packet. Therefore, we configure the iBeacon UUID such that the resulting transformed value can still represent the user's UUID, even after iOS modifies the data in the background.

When a single iBeacon UUID is set, one bit out of 128 bits changes from 0 to 1. As an example, in Fig. 4, the iBeacon UUID is 0466cb00-a9e0-4414-945c-43dd4eec4c6c. In this case, the 19th bit out of the 128 bits changes. The smartphone beacons advertise a value of 4c0001 for the fixed value portion and 00000000000000000000000080000 for the value that changes according to the iBeacon UUID. Similarly, when the iBeacon UUID 64a84cda-101f-43af-b871-259d37d4309a is set, the 58th bit out of the 128 bits changes. Therefore, the smartphone beacons advertise a value of 4c0001 for the fixed value portion and 00000000000000000400000000000000 for the value that changes according to the iBeacon UUID.

When multiple iBeacon UUIDs are set, multiple bits within the 128 bits change from 0 to 1. As an example, the two iBeacon UUIDs given in the previous example are set in Fig. 5. In this case, the 19th and 58th bits out of the 128 bits are changed. he smartphone beacon advertises a value of 4c0001 for the fixed value portion and 00000000000000000400000000000080000 for the value that changes according to the UUID. When multiple iBeacon UUIDs are set in this way, all the bit positions corresponding to each iBeacon UUID will change.

3.4.2 How to Transmit BLE Signals Using The Correspondence Table

We created a correspondence table to express the value to be advertised in the packet using the above properties. The table is shown in Fig. 6.

The smartphone beacon uses the table to set multiple iBeacon UUIDs to represent the values it wants advertised in the packet. The example is shown in Fig. 7. The smartphone beacon wants to advertise 0000000-00a0-0000-0000-0000-000000000005000. In this case, the value advertised should be 4c0001 for the fixed value part and 0000000000000000a0000000000000000000000000005000 for the value that changes according to the iBeacon UUID. To achieve this, because bits 12, 14, 84, 85, and 87 must be changed from 0 to 1, the iBeacon UUID corresponding to each bit position should be checked in the corresponding table. When all the corresponding iBeacon UUIDs are set and advertised, the smartphone beacon advertises the following value.
4c0001000000000000000000000000000000a000000000000000000000000000000000000000000000000000000000005000
This allows the UUID to be advertised from an iOS device like a physical beacon.

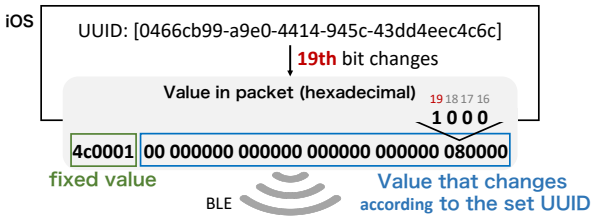


Figure 4: Example of transmitting BLE signals in the background on iOS

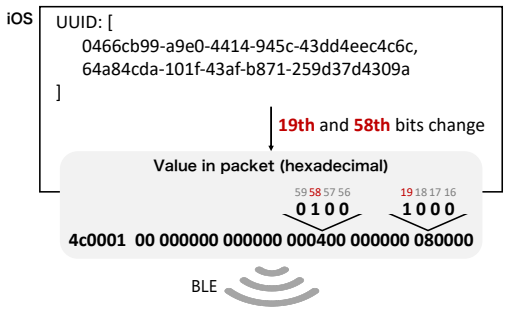


Figure 5: Example of multiple UUIDs being sent out in the background on iOS

Changing bit positions	UUID to be set
0	00000000-0000-0000-0000-000000000039
1	00000000-0000-0000-0000-000000000072
2	00000000-0000-0000-0000-00000000002b
126	00000000-0000-0000-0000-000000000037
127	00000000-0000-0000-0000-00000000007c

Figure 6: Table of correspondence between bit positions to be changed and iBeacon UUIDs to be set

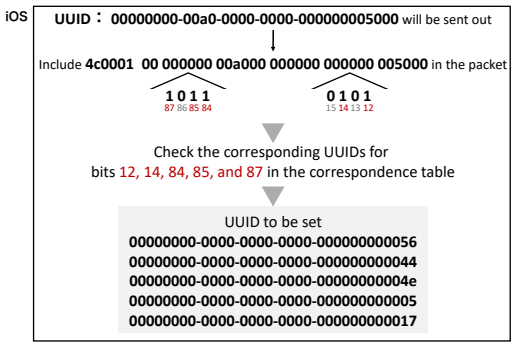


Figure 7: Example of setting multiple iBeacon UUIDs with correspondence table in iOS

However, even if the iBeacon UUID to be advertised is expressed in this way, it will conflict with other applications. The overflow area that contains the set iBeacon UUID is common to all applications. Therefore, the value in the packet may be affected by other applications that transmit BLE signals. Even if the user's UUID is contained in the packet, it may be changed by other applications. In this case, the user cannot be identified by the stay estimation system that identifies the user based on a perfect match of UUIDs.

To address this issue, a part of the user's UUID is expressed in multiple places in the packet. The administrator sets the UUID of the iOS user in advance to a UUID that can be identified if a part of the UUID is known. This allows the system to identify the user even if some bits are changed by other applications, as long as other parts are unaffected. In this way, the smartphone beacon satisfies **R1**.

3.4.3 How to Prevent BLE Radio Advertising from Stopping

We made the smartphone beacon automatically resume transmission when it becomes possible to transmit when it is unable to continue transmitting BLE signals to satisfy **R4**. There are two scenarios where BLE signal transmission cannot be continued in iOS: when the smartphone is turned off and when the application is forced to terminate by the system.

We used the state preservation and restoration feature to address these issues. The feature allows the system to take over the background process and start the application in the background when necessary, even if the application is stopped. This allows the system to request that BLE-related tasks be performed in place of the application when it is stopped. In this way, the smartphone beacon satisfies **R4**.

The smartphone beacon targets devices running iOS 6 or later because both CoreBluetooth and state preservation and restoration were introduced in iOS 6. We confirmed its correct operation on devices running iOS 17.0.3 through iOS 18.5.

4 EVALUATION EXPERIMENT

4.1 Tracking Comparison Experiment

We deployed the smartphone beacon on the stay estimation system to investigate whether the smartphone beacon has better tracking performance than a physical beacon. We conducted an experiment to compare the actual number of stays with those detected by smartphone beacons and those detected by physical beacons.

4.1.1 Experimental Setup

The subjects were 13 university students belonging to the same laboratory. The OS used was Android for subjects 1-6 and iOS for subjects 7-13. The experimental period was from October 10, 2023 to December 26, 2023. However, for iOS subjects, the application built directly from a PC had a validity period of only one week, after which the application would no longer function. Although the application was periodically rebuilt and redistributed, some subjects experienced expiration before reinstallation. Therefore, the experimental period for iOS subjects was limited to the application's validity period. Subjects will be asked to install the smartphone beacon on their own smartphones, and both the smartphone beacon and the physical beacon were kept running during the experiment period. Figure 8 shows an image of the room used for the experiment. One receiver was placed on a desk in the center of the room. A camera was installed at an angle at the room's entrance and exit that allows identification of people entering and exiting. The camera recorded continuously during the experiment period to collect actual stay information. The correct answer rate was defined as the percentage of correctly detected stays out of the actual number of stays.

4.1.2 Experimental Results

Table 1 shows the results of the evaluation experiment. In the table, the correct answer rate for the smartphone beacons is defined as CAR_S and that for the physical beacons as CAR_P. The average correct answer rate for each user was 85% for the smartphone beacon and 62% for the physical beacon. Eleven out of thirteen subjects had a higher correct answer rate for the smartphone beacon. This result suggests that the smartphone beacon has a higher tracking performance than the physical beacon. This may be due to the fact that some subjects left the physical beacon in their room moving between rooms for classes or club activities, whereas many subjects tend to carry their smartphone even when moving between rooms. While some subjects using smartphone beacon

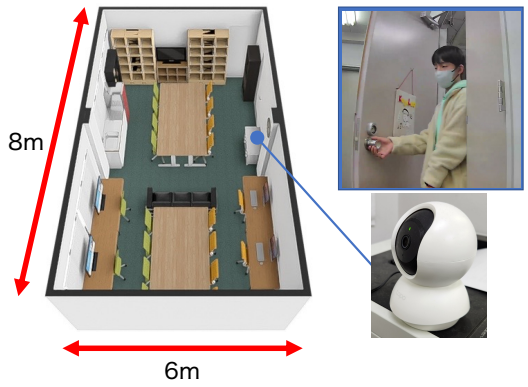


Figure 8: Room used for experiments. A camera records the entrance and exit at all times.

Table 1: Comparison of tracking performance by subject

Subject	Number of stays (times)	CAR_S (%)	CAR_P (%)
1	93	98	92
2	50	96	92
3	36	100	36
4	32	75	78
5	44	89	82
6	40	83	38
7	98	79	60
8	75	97	32
9	48	90	54
10	24	75	50
11	58	100	50
12	24	88	63
13	20	80	90
avg.	–	**88**	**63**

had a correct answer rate of over 90%, others had a correct answer rate of less than 80%.

In order to prevent unexpected behavior in Android, the application stops advertising BLE signals when Bluetooth is turned off. Some subjects frequently turned off the Bluetooth function of their smartphones in order to connect their Bluetooth devices to other devices instead of their smartphones. This was the main reason why the correct answer rate decreased in Android. However, Android can continue to transmit BLE signals even when Bluetooth is turned off. Therefore, it is expected that modifying the implementation will lead to a higher correct answer rate for the smartphone beacon.

4.2 Battery Consumption Comparison Experiment

The smartphone beacon suppresses battery consumption, and we conducted an experiment to compare the battery consumption to evaluate whether the beacon was able to satisfy the **R2** defined in 4.1. In order to compare this method with a method in which a physical beacon is placed in the environment and received by a user's smartphone, we implemented an application that receives surrounding BLE signals (hereafter referred to as "scanning application").

Table 2: 24-hour battery consumption per device

Device (ad interval, receive interval)	None	Beacon	Scan
Galaxy S10 (130～140ms, 3s)	8%	9%	42%
Pixel 7 (130～140ms, 10s)	6%	6%	18%
Xperia 10 V (130～140ms, 10s)	2%	2%	5%
iPhone 11 Pro (270～280ms, 1s)	0%	1%	3%

4.2.1 Experiment Setup

The four smartphones used in the experiment are Galaxy S10, Pixel 7, Xperia 10 V, and iPhone 11 Pro. They are left for 24 hours in each of the following conditions: no activity, with the smartphone beacon running, and with the scanning application running. Each smartphone is charged to 100% beforehand. To prevent external factors, no network connection is made. After 24 hours, we measured the battery consumption. The transmission and reception intervals of the BLE signals are not strictly set by the developer. The transmission interval of the smartphone beacon is 130-140 ms for Galaxy S10, Pixel 7, and Xperia 10 V, and 270-280 ms for iPhone 11 Pro. The receiving interval of the scanning application is 3s for the Galaxy S10, 10s for the Pixel 7 and Xperia 10 V, and 1s for the iPhone 11 Pro.

4.2.2 Experiment Results

Table 2 shows the results of the evaluation experiments. Although there are differences depending on the device, in all cases, the smartphone beacon that periodically transmits BLE signals consumes less battery power than the scanning application that continuously receives BLE signals. There is almost no difference in battery consumption between running the beacon and having no application running. This is thought to be due to the fact that transmitting BLE signals consumes little power, and that smartphone beacons are single function applications that only advertise BLE signals.

Pixel 7 and Xperia 10 V consume about three times as much battery power when the scanning application is running as when no application is running. The Galaxy S10 consumes about five times as much. This may be due to the fact that the interval between scans is shorter than that of other devices.

The scanning application is intended for applications that use smartphones to receive BLE signals from physical beacons placed in the environment. This method is expected to consume more battery power than the scanning application used in this experiment, since it also requires periodic transmission of the received physical beacon information to the server. Therefore, the smartphone beacon is considered to be an application with low battery consumption.

5 CONCLUSION

In this paper, we discuss the issues of using only physical beacons in a stay estimation system using BLE, and implement and evaluate a smartphone beacon that enables smartphones to behave like a physical beacon. The method using only physical beacons has problems such as high financial cost, time and effort for initial setup, and forgetting to replace batteries. We have developed a smartphone beacon with high tracking performance that does not require initial setup or battery replacement. However, because smartphone beacons also have disadvantages, we made it possible for the user to choose which beacon to use. In the evaluation experiment, we actually installed the smartphone beacons in a stay estimation system and investigated the tracking performance and battery consumption of the smartphone beacons. The results showed that smartphone beacons had higher tracking performance and lower battery consumption than physical beacons.

Current the smartphone beacons are always traceable by unintended third parties because the UUIDs included in the BLE advertising packets are fixed and unchanging for each user. The transmitted BLE signals can be received by electronic devices such as BLE-compatible smartphones, and the packets can be easily browsed. Therefore, as long as the user's UUID is known, a third party can determine when that user was near a receiver by installing a receiver or other device without permission. We believe that a solution to this problem requires a mechanism that allows the stay estimation system to identify whose BLE beacon terminal the signal came from, but prevents third parties from determining the owner of the BLE beacon terminal.

Currently, we only use received radio wave strength for estimation, which may cause misjudgments when the user is in an adjacent room, depending on the material and thickness of the wall. If receivers are installed in adjacent rooms, it would be possible to compare the reception strength of each receiver for more accurate room-level localization. Therefore, it is necessary to deploy and validate our stay estimation system in other communities.

REFERENCES

[1] C. J. Matz, D. M. Stieb, K. Davis, M. Egyed, A. Rose, B. Chou, and O. Brion, "Effects of age, season, gender and urban-rural status on time-activity: Canadian-Human Activity Pattern Survey 2 (CHAPS 2)." *International journal of environmental research and public health*, Vol. 11, No. 2, pp. 2108-2124 (2014).

[2] ACCESS CO., LTD. Linkit Kintai, https://linkit.access-company.com/kintai/ (reference 2024-11-14).

[3] UCHIDA YOKO CO., LTD. SmartOffice-Navigator, https://www.uchida.co.jp/it/products/smartofficenavigator/ (reference 2024-11-14).

[4] K. Suda, T. Koinuma, and T. Suzuki, "Attendance Registration System Using Contactless IC Cards." *IEICE Technical Report*, Vol. 111, No. 141, pp. 65-70 (2011).

[5] K. Ishizawa, and M. Iwai, "A System of Managing Person Staying in Room Based on Utterance and Recognition Area." *UBI Technical Report*, Vol. 57, No. 19, pp. 1-6 (2018).

[6] Bitkey Inc. workhub, https://www.bitlock.workhub.site/product/face-recognition. (reference 2024-11-14).

[7] A.B.M Musa, and J. Eriksson, "Tracking Unmodified Smartphones Using Wi-fi Monitors." *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, pp. 281-294 (2012).

[8] R. Faragher, and R. Harle, "Location Fingerprinting With Bluetooth Low Energy Beacons." *IEEE Journal on Selected Areas in Communications*, Vol. 33, No. 11, pp. 2418-2428 (2015).

[9] P. Barsocchi, A. Crivello, M. Girolami, F. Mavilia, and F. Palumbo, "Occupancy detection by multi-power bluetooth low energy beaconing." *2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1-6 (2017).

[10] A. Tokuda, Y. Arakawa, S. Takano, and S. Ishida. "Examination of Automatic Parameter Adjustment in Hybrid Congestion Measurement by WiFi and BLE." *DPS Technical Report*, Vol. 187, No. 16, pp. 1-8 (2021).

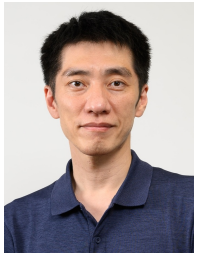
[11] Labour Ministry of Health and Welfare. COVID-19 Contact-Confirming Application (COCOA), (2020) https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/cocoa_00138.html, (reference 2024-11-14).

[12] Cabinet Office. Smart City, https://www8.cao.go.jp/cstp/society5_0/smartcity/, (reference 2024-11-14).

[13] Y. Tanaka, T. Fukushima, and T. Yoshino, "Docoitter : A Presence Display System Capable of Predicting Future In-the-room Information." *IPSJ Journal*, Vol. 54, No. 9, pp. 2265-2275, (2013).

(Received November 15, 2024)
(Accepted August 15, 2025)

Kota Togawa is a graduate student at Aichi Institute of Technology. Currently, he is a graduate student of Business Administration and Computer Science, Aichi Institute of Technology. His research interests include indoor positioning.



Katsuhiko Kaji received his Ph.D.in information science from Nagoya University in 2007. He be-came a RA at NTT Communication Science Lab-oratories in 2007 and an assistant professor in Nagoya University in 2010. He moved to Aichi Institute of Technology in 2015 as an associate professor, be-coming a professor in 2024. His research interests include indoor positioning, human activity recog-nition, and human augmentation. He is a member of IPSJ.

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