Efficient Object Retrieval System Using Contact History from a Thermal Camera

Mone Tanaka*, Yasue Kishino†*, Hideyuki Shimonishi‡
*Graduate School of Information Science and Technology, Osaka University, Japan
tanaka.mone@ist.osaka-u.ac.jp
†NTT Communication Science Laboratories, Japan
‡D3 Center, Osaka University, Japan

***Abstract*—A considerable amount of time and effort is expended in searching for lost objects. Thus, various systems have been developed to assist in this endeavor. For example, radio-frequency identification (RFID) tags are useful in indoor environments but are inefficient due to the battery lifetime and the cost of attaching them to many objects. Camera-based systems can track many objects at once, but if they track all objects that appear in the user's surroundings, a lot of candidates are suggested, and the selection requires a great deal of effort. To solve this problem, we propose a system for finding lost items using RGB and thermal cameras. The proposed system combines object detection using the RGB camera and contact detection using the thermal camera to obtain history information limited to objects touched by a person. User study shows that our proposed system reduced the object retrieval time by 47.3% and user workload by 32.7% compared with the comparison system. Furthermore, participants evaluated our system as useful for object retrieval according to the System Usability Score (SUS).**

***Index Terms*—Lost Object Finding, Memory Aid, Contact Detection, Thermal Cameras.**

I. Introduction

Everyone has experienced losing something essential, such as keys, wallets, or glasses. A recent survey found that people lose things in their homes on average 3.2 times a month[1]. Another survey found that people spend 2.5 days a year searching for lost items[2]. Therefore, support systems for searching for lost items have great potential to make people's lives more convenient.

Various conventional systems are available to search for lost objects [1], [2]. Attaching wireless tags [3] or augmented reality (AR) markers [4] is a typical approach. Although these techniques can reliably identify an object's position, tracking all objects is difficult because of the time-consuming attachment process and the cost of tagging multiple items.

According to a survey conducted by Yan et al. on 120 people, the most frequently cited reasons for losing items in public places such as offices were "because the objects have been moved by others" (58 people), "because the objects do not catch people's attention" (51 people), and "because the objects were taken by me but forgotten" (41 people).

Thus, we can infer that objects are often lost when moved. Also, it is assumed that less important objects that do not have loss-prevention wireless tags attached are more likely to be lost. Therefore, a camera-based system is more effective than a wireless-tag-based system in searching for lost items. However, if a camera-based system tracks all objects around the user, the number of candidate objects becomes huge.

[1]https://prtimes.jp/main/html/rd/p/000000211.000022173.html

[2]https://web.archive.org/web/20171206223019/https://getpixie.com/blogs/news/lostfoundsurvey

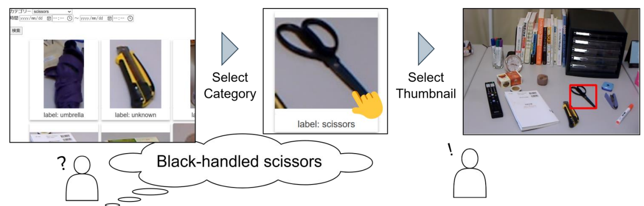

Fig. 1: Image of the proposed system. Users can narrow their search by category or time.

To reduce the number of candidates, Yagi et al. [5] focused on the fact that most lost items are handheld objects. Their system used a wearable camera and limited search items to hand-held objects, rather than all objects. However, this system cannot respond to multi-person scenarios because it only captures images from a first-person viewpoint. In a production environment, multiple people often share the object in the same space, such as in an office or at home. Thus, in this study, we use a fixed camera to deal with these situations. In addition, we narrow down the search scope efficiently to reduce the user's burden by introducing a thermal camera.

Thermal cameras measure heat on the surface of an environment and can detect the point of contact where a person touches an object. Using the information from the detected contact point, our system addresses only the contacted objects. In this paper, we describe our proposed system that supports users' finding lost objects using RGB thermal cameras.

II. Related Works

There has been a lot of research on lost item search, and various methods have been proposed to assist, including methods that use wireless tags, fixed cameras, and wearable cameras. Radio-frequency identification (RFID) tags are effective when used in indoor environments, but they cannot locate an object outside the search range. For a wider search range, commercially available products employ a system that combines Bluetooth and GPS. These wireless-tag-based systems can provide the location of objects, but tagging all objects is impractical. Camera-based systems, on the other hand, do not require wireless tags and have the advantage of being able to track a large number of objects.

Ueoka et al. [6] proposed the "I'm here!" system that uses a head-mounted device equipped with a visible camera and an infrared camera to recognize and record objects that the user has registered in advance. The system displays a video of the last time the object was captured on the head-mounted display to assist the user in searching for the lost object.

Noting that most lost objects are handled by hand, Yagi et al. proposed "GO-Finder" [5], which uses a wearable camera to assist in the retrieval of objects grasped in the past. GO-Finder provides the user with the last image of the tracked object to assist in finding it. It can track an arbitrary object by using the image of the object as a search query without registering it beforehand. The user study confirmed that GO-Finder effectively reduces psychological burden compared to a situation without assistance. However, because GO-Finder collects and logs images from a first-person video, it does not support situations where multiple people share objects.

Yan et al. [7] proposed "CamFi," which combines face recognition and object detection to find lost items in multi-user scenarios. CamFi detects objects with a fixed camera and provides information on the person who last touched the object or who was captured most often at the same time as the object. In user experiments, 9 out of 12 participants found the system helpful, but some privacy concerns were raised.

III. Contact Detection

Normally, the human body temperature is higher than the environmental surface temperature. When a person touches an object, they leave a heat trace; the heat from the person's hand remains on the object's surface. In Fig.2, heat traces remain on the calculator even after the person's hand is removed.

Research using heat traces has been conducted in the past, especially in the field of interaction, and attempts have been made to use heat traces as an input interface [8], [9].

Kishino et al. [10] used a thermal camera to visualize human touch points to prevent infectious diseases. This study implemented touch detection by detecting heat traces using background subtraction image processing. The study used the static state of the location in front of the fixed camera as the background image, and human contact was detected by applying processing to the difference between the background image and the current image.

In our study, we used this technology to limit the search range from many objects captured by an RGB camera to objects touched only by people. When a thermal camera detects heat traces and an RGB camera simultaneously detects changes on the same object, we can assume that the object has been moved. We can therefore expect the search range for lost items to be narrowed down significantly, making information on lost items more easily accessible.

IV. Proposed System

As mentioned in Chapter 1, losing things in a shared space is often triggered by moving objects. Since most objects are moved by hand, the history of objects touched by people can support the search for lost objects. Therefore, we focused on contact detection using thermal cameras. Figure3 shows the structure of the proposed system. This study assumes an indoor space shared by several to a dozen people.

Fig. 2: Example of a heat trace

TABLE I: Object detection and contact detection relationship

Object detection	No change in thermal image	Change detected in thermal image
Start	Exist continuously	**Placement or used (Detection of heat traces)**
Stop	Occlusion	**Carry-away** (or occlude by people)

The proposed system first detects objects people touch using RGB and thermal cameras, and records contact object logs in a database. Next, the contact object logs are summarized as a contact history. The system shows the contact history information on a web page and the user can check a photo of the last time an object was used or touched. The Web page contains a list of the following information, based on previous studies [5], [7], a thumbnail of the object only, an image of the last time the object was seen, and other detailed information. When searching for a lost object, this information makes it easy for users to know when and where the lost object was placed by looking at the image captured the last time it was touched. This is especially effective in cases where an object is missing in a shared space.

Our proposed system is advantageous compared to previously proposed systems because it uses the presence or absence of changes in the thermal image as a search index. Table I summarizes the relationship between object detection and contact detection. If there is no change in the thermal image, we assume the object remains in place and has not moved. On the other hand, if there is a change in the thermal image, we assume that the object has moved. Therefore, we focus on the presence or absence of a change in the thermal image and assume that the system can capture the movement of an object that triggers its loss. That means we can use the change as a search index.

A. Contact Object Detection

Our system detects contacted objects using images captured by RGB and thermal cameras. First, we detect objects using YOLOv8 [11] and then list them. Then, heat trace detection is performed on the thermal image. When the detected heat trace is inside the bounding box of the object detected in the RGB image, it is recorded as a contact object log. Heat traces are extracted through background subtraction and binarization, after which areas with temperature differences above a certain

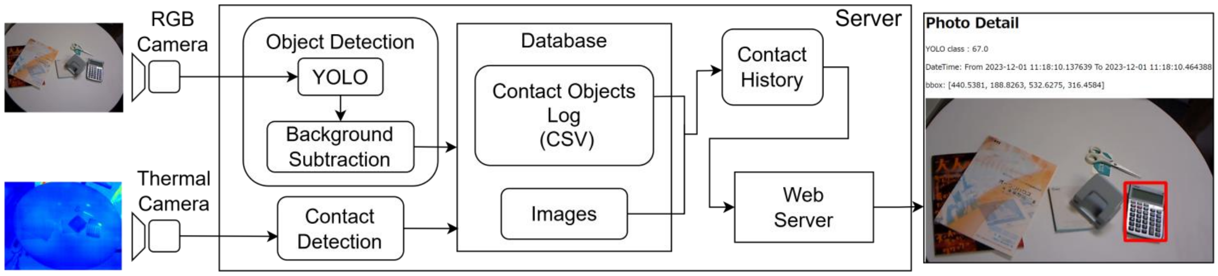

Fig. 3: Configuration of the proposed system

threshold are identified. The centroid is calculated from the contour of the extracted heat traces. If the centroid of the heat trace is inside the bounding box, contact with the object is considered to have occurred.

The contact object log is recorded with the YOLO class label, detection time, and bounding box coordinates. At the same time, the entire frame image and the RGB and thermal thumbnail image of only the object cropped by the bounding box are stored in the database.

B. Integration from Contact Object Log to Contact History

Next, we describe the method for converting the contact object logs collected above into a contact history. In the previous steps, objects touched by people are automatically detected and collected as search candidates. However, because each is an independent log, it is necessary to integrate them as a series of contact histories. This step includes determining whether the object recorded in the log is the same as an object that has been previously registered. In this operation, we use the histogram of the hue channel in the HSV (hue, saturation, and value) color space and the IoU (Intersection over Union) of the bounding box. We use hue because its value is independent of brightness and is less affected by lighting. Suppose the sum of the comparison result of the histogram and the IoU value of the bounding box is greater than the threshold value. In that case, the object is considered identical to a previously registered object. When calculating the histogram values, a mask image is generated by a thermal camera to remove any human hand parts to avoid misrecognition of the image in the process of being touched, and the RGB image is masked to calculate the histogram of the object only. This procedure is performed in the following 3 steps.

1. If the object label in the contact object log is not found in the record, it is registered as a new contact history.
2. If the label has already been registered, the IoU of the histogram and the bounding box are used to determine whether the object is identical to a previously registered object. If it is determined to be a different object, it is registered as a new contact history. If determined to be the same object, the contact log is considered to be a part of the corresponding contact history information, and the average value of the bounding box and the last confirmed time in the contact history are updated.
3. When a certain amount of time (e.g. 30 sec in this implementation) has passed since the last object was confirmed, the last confirmation time and the average value of the bounding box of the contact history information are determined. If the number of frames in which an object is detected is less than a threshold value (25 frames in this implementation), it is not recorded because there is a high possibility of false detection.

These processes store the contact history information of the objects touched by people within the camera's field of view.

C. Search Interface

If the objects to be searched for are managed by name or ID, the user is burdened with the time and effort of pre-registration. Therefore, we adopted a search interface that displays a list of thumbnail images so that users can quickly find objects without registering them in advance. Also, the location of the last time an object was placed is important for the user to find lost items. For these reasons, the user can select the image of the lost object from the thumbnail images to see the image of the last time the object was captured.

The implemented search interface is shown in Fig.1. This search interface presents information such as pictures of objects that the user has touched so far by using the contact history information made in the previous sections. The user can narrow down the search target by category, date, or time. When the user clicks on the object thumbnail to search for an object, detailed information is displayed, including the image of the last time the object was seen, the YOLO label, the recorded time, and the bounding box coordinates. In the current implementation, objects detected by background subtraction processing without YOLO are labeled as unknown. Although they are treated as unknown categories in the current implementation, we are considering using specific category names for these objects in the future. The object is indicated by a red box around it on the image when the object is last captured. The red box is drawn based on the average of multiple bounding boxes stored in the log.

V. System Evaluation

We conducted the following experiments to verify the effectiveness of introducing a thermal camera to support the search for lost objects.

Fig. 4: Experimental setup (Left: Thermal and RGB Cameras used in the experiment, Right: Scene of the experiment).

TABLE II: Results of Experiment 1

	YOLO	YOLO Contact Detection
Success	11	10
Detection Leak	0	1
Failure (recognized multiple times)	5	0

A. Experimental Environment

We used an RGB camera (GAZO MCM-303NIR) with a resolution of 480×640 and a thermal camera (FLIR BOSON640R) with a resolution of 480×512 for the experiments. Figure4 shows the experimental setup. The RGB and thermal images were aligned using an affine transformation, and then the unnecessary parts of the thermal image were cropped to match the coordinates and size of the RGB camera.

In Experiment 1, we prepared 9 types of objects, 11 in total, that can be detected in the YOLO class labels: books, scissors, calculators, smartphones, earphone cases, cloth pencil cases, water bottles, plastic bottles, and cosmetic bottles. Each object was placed on the table, moved, and removed repeatedly, and we videotaped it to verify the image processing results. The video was approximately 3 minutes 30 seconds long, with 6070 frames and a frame rate of 30 fps.

In Experiment 2, we used 27 objects of 13 types. The objects included in the YOLO class labels were remotes, scissors, plastic bottles, books, magazines, cups, smartphones and earphone cases. The objects not included in the YOLO class labels were keys, wallets, pens, notepads, and documents. As in Experiment 1, each object was repeatedly placed on the desk, moved, and removed, and the images taken were used to verify the results. This video was about 3 minutes long, with 5849 frames and a frame rate of 30 fps.

B. Results

1) Results of Experiment 1: The results of Experiment 1 are shown in Table II. We compared the results of the proposed method using contact detection with those using only YOLO.

When only YOLO was used, the thumbnail image of the same object was displayed 5 times. This may be because when an object that had been occluded is detected again, it is registered as a newly appeared object.

2) Results of Experiment 2: Table III shows the experimental results, and Fig.5 shows examples of recognized images. Using background subtraction, we detected 7 objects from 16 objects that were not detected by YOLO. The objects that could not be detected by YOLO or background subtraction

TABLE III: Results of Experiment 2. The numbers in parentheses indicate the number of candidates that contact detection was successful. However, problems were encountered, such as the same object being detected separately or the same object being detected by both YOLO and background subtraction.

(a) YOLO + Contact Detection

		Actual	
		Contact	No-Contact
Predicted	Contact	14 (5)	0
	No-Contact	16	N/A

(b)YOLO + Background Subtraction + Contact Detection

		Actual	
		Contact	No-Contact
Predicted	Contact	21 (10)	9
	No-Contact	9	N/A

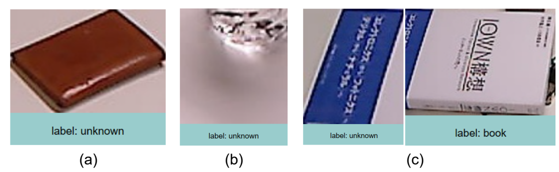

Fig. 5: Examples of object detection by background subtraction. (a) Successful object detection. (b) Failure of object detection by background subtraction. (c) Two examples of the same object detected separately.

were those with small differences from the background, such as a white notepad or a document on a white desk.

On the other hand, 9 false positives occurred, which is possibly due to reflection or differences caused by shadows cast by objects or people. In some cases, the same object was displayed separately. In other cases, objects that were already detected by YOLO were detected again by background subtraction. Although the area of the object detected by YOLO was excluded from object detection by background subtraction, this may have been due to insufficient detection of the object by YOLO or detection of only a narrower area than the original area of the object.

In conclusion, although there is room for improvement, we confirmed that by introducing background subtraction, the proposed system can address objects that are not in the YOLO class as candidates.

VI. User Study

We conducted a user study to evaluate the system and validate the hypothesis that we can find lost objects more effectively using the proposed system than a system that records all objects that appear in the camera's angle of view. Our user study was approved by Osaka University's ethics committee with which the authors are affiliated. We assume a certain individual who knows how to use this system has 3 desks of different purposes in various locations and loses things. Participants were asked to retrieve a target object under

Fig. 6: The experimental setup of the table.

the conditions of being assisted by the proposed system or compared method.

A. Participants and Experimental Conditions

22 individuals (11 males, 11 females) aged 19 to 25 years participated in the user study. We excluded one female participant due to a lack of data. All were familiar with the use of laptops.

In the experimental setup, we prepared 3 identical tables and 70 objects (Fig.6). Each object was placed on the table and video was captured simulating a few days of the object moving likewise the system evaluation experiment procedure. We used 9 minutes of the video for thumbnail extraction and contact history image processing. We used the same cameras for the system evaluation experiment. In the user study, we compared the proposed system and the system without a thermal camera. After image processing of the captured video, the number of thumbnail images was 35 for the proposed method and 111 for the compared method.

B. Procedure

Each participant was informed about the purpose of the experiment. We then conducted the 2 following experiments.

a) Experiment 1: In Experiment 1, we explained to the participants how to use the system and directed them to find a target object on the user interface (UI). After the trial was repeated 3 times for 3 different objects, participants answered the NASA Task Load Index (NASA-TLX) within a 100-point range with 5-point steps and provided feedback on the task. We aimed to evaluate how contact history increases the system's efficiency. We adopted the Raw Task Load Index (RTLX) as a simple way to measure the workload [12]. Then we changed the experimental condition and conducted the same trials with 3 target objects. The experimental conditions were randomly shuffled to eliminate the effect of order.

b) Experiment 2: In Experiment 2, we moved to the experimental room and asked the participants to find a target object on the table using the UI and retrieve it. After each trial, the participants answered the 5-point System Usability Scale (SUS) questionnaire [13] and provided feedback on the task. We aimed to evaluate whether the system works well for object retrieval. Thus, each participant was asked to search for a target object using the UI. We randomly shuffled the experimental conditions.

C. Results

1) Experiment 1: In Experiment 1, we conducted an objects-retrieval task on a UI to evaluate how the contact history acquired by a thermal camera works for system usability.

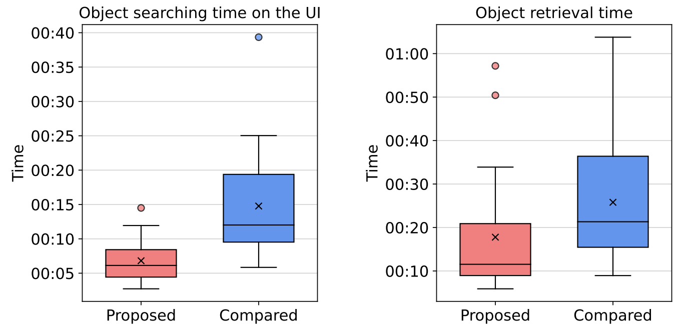

Fig. 7: Left: Average time to find a target object on the UI, Right: Object retrieval time

a) Time: The average time among the 3 trials to find a target object on the UI is shown in Fig.7 We used the Wilcoxon signed-rank test to investigate whether there was a significant difference in the average time among the 3 trials to find a target object in the UI between experimental conditions. The average time to find a target object on the UI is $6.56 Sec$ ($\sigma = 3.00$) for the proposed system and $13.55 Sec$ ($\sigma = 6.03$) for the compared system. The results show that the participants take significantly less time when assisted by the proposed system ($W = 0$, $\alpha = 0.05$, $p = 1.91 \times 10^{-6}$).

b) NASA-TLX: The participants answered the NASA-TLX within a 100-point range with 5-point steps. 2 participants who likely misunderstood the intent of the questions were excluded from the evaluation. The average NASA-TLX score was 116.19 ($\sigma = 71.17$) for the proposed system and 172.62 ($\sigma = 110.59$) for the compared system. The Wilcoxon signed-rank test shows a significantly lower workload for the proposed system ($W = 34$, $\alpha = 0.05$, $p = 2.48 \times 10^{-2}$). There is no difference between these two systems, aside from the number of thumbnail images, so it appears that the image count affected the object searching time on the UI and the NASA-TLX score. It could be time-consuming and stressful for participants to find a target object among many images.

2) Experiment 2: In Experiment 2, we conducted an object retrieval task to evaluate whether our proposed system is useful for searching for lost objects.

a) Object Retrieval Time: Object retrieval time is shown in Fig.7. We compared the object retrieval time among the different conditions with the Wilcoxon signed-rank test. As a result, the average object retrieval time is $14.02 sec (\sigma = 7.31)$ for the proposed system and $26.60 sec (\sigma = 16.40)$ for the compared system. The proposed system significantly saved time compared to the comparison system ($W = 38$, $\alpha = 0.05$, $p = 3.85 \times 10^{-2}$). It is supposed that the number of thumbnail images affected the object searching time on the UI, resulting in a significant difference in object retrieval time.

b) SUS score: For an average SUS score, our proposed system received 75.88 ($\sigma = 11.25$), and the compared system received 72.63 ($\sigma = 15.44$). There is no significant difference in the SUS score between the proposed system and the compared system, which is likely because there are no other differences in the UI of the two systems, except for the number of thumbnail images. However, according to a previous study [14], a score of 70 or higher is considered acceptable. Therefore, it can be said that participants found the proposed method useful for searching for lost items.

VII. Discussion

This study explored object retrieval using RGB and thermal cameras, revealing that our system reduced object retrieval time and user workload. This discussion will further analyze these results and examine potential avenues for future research.

A. Usefulness of the Proposed System

In this paper, we propose a lost object-finding system using a fixed RGB camera and a thermal camera that limits the candidates to objects touched by a person. The results of the user study indicate that our system reduces object retrieval times and perceived workload. Although the SUS result has no statistically significant difference between the proposed and comparison systems, the NASA-TLX results indicate that our proposed system effectively reduces user workload. This system has the potential for practical use in shared spaces with some people by setting up multiple cameras.

Regarding our user study design, 22 participants were valid enough based on the variability of the evaluation results. 5 participants answered that they had difficulty finding the target object when the angle of the object displayed on the UI was different from the one specified in the photo. Participants were asked to search for unfamiliar objects; however, assuming that the system would be used in daily life, users would search for familiar objects. Thus, we believe that this is not a critical issue but might affect the NASA-TLX and the SUS score.

Also, in this experimental setup, the 3 tables were side by side, so participants could find objects without much movement. On the other hand, in our daily lives, the places we place objects are not necessarily side by side. Our system is very useful, especially in situations such as when you want to know whether something you have lost is at home or office.

B. Limitations and Future Work

The fixed camera approach, like the proposed method, is expected to detect when someone touches and moves an object for all users of a shared space and to be useful in searching for lost items compared to the wearable camera approach. On the other hand, it cannot detect object contact with occlusions or outside the camera's angle of view. This approach is effective in situations where shared space is limited, such as on desks or shelves, but multiple cameras are required in situations where the entire room has to be covered.

In the user study, 8 participants answered that categories are too broad and confusing. This must be because all objects not detected by YOLO are categorized as unknown. This will be resolved as users become accustomed to this system, or fine-tuning the YOLO. Also, in the user study, we asked participants to search for unfamiliar objects that we had placed. Therefore, the task may have been more difficult than in the case in which users search for what they use every day. In future research, we would like to conduct experiments that more closely resemble the real environment by asking users to search for familiar objects placed by themselves.

VIII. Conclusion

People spend a lot of time and effort searching for lost objects. To solve this problem, we propose an object retrieval support system using RGB and thermal cameras. The proposed system uses the RGB camera for object detection and the thermal camera for contact detection and stores historical information about objects the user has touched. This alleviates the high cognitive burden on the UI of recording all objects that appear inside the camera's angle of view. The user study revealed that by utilizing the contact information from the thermal camera, users could search for target objects more efficiently than with the system that records all the objects. In addition, the proposed system reduced participants' perceived workload and was evaluated as a useful system for object retrieval. In future work, we plan to conduct a user study closer to a real environment.

References

[1] R. Fukui, T. Sunakawa, S. Kousaka, M. Watanabe, T. Sato, and M. Shimosaka, "Tansubot: A drawer-type storage system for supporting object search with contents' photos and usage histories," in *International Conference on Intelligent Robots and Systems*. IEEE, 2013.

[2] H. Oshimi, M. Perusquía-Hernández, N. Isoyama, H. Uchiyama, and K. Kiyokawa, "Locatar: An ar object search assistance system for a shared space," in *Proc. of the Augmented Humans International Conference*, 2023.

[3] S. M. Huynh, D. Parry, A. Fong, and J. Tang, "Novel rfid and ontology based home localization system for misplaced objects," *IEEE Transactions on Consumer Electronics*, vol. 60, no. 3, 2014.

[4] A. Butz, M. Schneider, and M. Spassova, "Searchlight–a lightweight search function for pervasive environments," in *Pervasive Computing: Second International Conference, PERVASIVE*, 2004.

[5] T. Yagi, T. Nishiyasu, K. Kawasaki, M. Matsuki, and Y. Sato, "Go-finder: a registration-free wearable system for assisting users in finding lost objects via hand-held object discovery," in *26th International Conference on Intelligent User Interfaces*, 2021.

[6] T. Ueoka, T. Kawamura, Y. Kono, and M. Kidode, "I'm here!: A wearable object remembrance support system," in *Human-Computer Interaction with Mobile Devices and Services: 5th International Symposium, Mobile HCI*, 2003.

[7] G. Yan, C. Zhang, J. Wang, Z. Xu, J. Liu, J. Nie, F. Ying, and C. Yao, "Camfi: An ai-driven and camera-based system for assisting users in finding lost objects in multi-person scenarios," in *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 2022.

[8] E. Larson, G. Cohn, S. Gupta, X. Ren, B. Harrison, D. Fox, and S. Patel, "Heatwave: thermal imaging for surface user interaction," in *Proc. of the Conference on Human Factors in Computing Systems*, 2011.

[9] D. Iwai and K. Sato, "Document search support by making physical documents transparent in projection-based mixed reality," *Virtual reality*, vol. 15, 2011.

[10] Y. Kishino, Y. Shirai, Y. Yanagisawa, K. Ohara, S. Mizutani, and T. Suyama, "Identifying human contact points on environmental surfaces using heat traces to support disinfect activities: poster abstract," in *Proc. of the Conference on Embedded Networked Sensor Systems*, 2020.

[11] G. Jocher, A. Chaurasia, and J. Qiu, "YOLO by Ultralytics," https://github.com/ultralytics/ultralytics, Jan. 2023, accessed on Dec. 17, 2023.

[12] J. C. Byers, A. Bittner, and S. G. Hill, "Traditional and raw task load index (tlx) correlations: Are paired comparisons necessary," *Advances in industrial ergonomics and safety*, vol. 1, 1989.

[13] J. Brooke *et al.*, "Sus-a quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, 1996.

[14] A. Bangor, P. Kortum, and J. Miller, "Determining what individual sus scores mean: Adding an adjective rating scale," *Journal of usability studies*, vol. 4, no. 3, 2009.