CamFi: An AI-driven and Camera-based System for Assisting Users in Finding Lost Objects in Multi-Person Scenarios

Ge Yan
Zhejiang University, China
yange@zju.edu.cn

Chao Zhang
Zhejiang University, China
zhangchaodesign@zju.edu.cn

Jiadi Wang
Zhejiang University, China
wjd1024@zju.edu.cn

Zheng Xu
Zhejiang University, China
22021297@zju.edu.cn

Jianhui Liu
Zhejiang University, China
greyljh@zju.edu.cn

Jintao Nie
Zhejiang University, China
niejt@zju.edu.cn

Fangtian Ying
Hubei University of Technology, China
13067848819@163.com

Cheng Yao
Zhejiang University, China
yaoch@zju.edu.cn

ABSTRACT

It is important to study how to help people quickly find misplaced objects. However, previous studies have focused on single-person scenarios without considering the influence of other people in public places. Based on the technology of object detection and face recognition, our system can help reduce the burden upon people's memory. It can provide useful information, whether the user forgets where the object is or because someone else has moved the object. The system includes a camera, processing server and smartphone application. To evaluate our approach, we conducted a quantitative and qualitative user study with participants (n=12). We demonstrated the usability of this system in helping users find misplaced items in public settings with multiple people.

CCS CONCEPTS

• **Human-centered computing**; • **Human computer interaction (HCI)**; • **Interactive systems and tools**;

KEYWORDS

Memory aid, Camera-based system, Object discovery, Public spaces, Object detection, Face recognition

ACM Reference Format:

Ge Yan, Chao Zhang, Jiadi Wang, Zheng Xu, Jianhui Liu, Jintao Nie, Fangtian Ying, and Cheng Yao. 2022. 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 (CHI '22 Extended Abstracts), April 29–May 05, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3491101.3519780

CHI '22 Extended Abstracts, April 29–May 05, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9156-6/22/04…$15.00
https://doi.org/10.1145/3491101.3519780

1 INTRODUCTION

It is common for people to forget where things are, but the impact in terms of time wasted may be surprising. A survey reported that it took people two-and-a-half days a year to look for misplaced objects [17]. Technological support can help users find lost objects and save time [8, 16].

Butz et al. [4] developed a stationary camera to check for things out of place, but it was limited as it needed to be set up to recognize the specific objects in advance. Xie et al. [22] developed a dual-camera wireless sensor network for object searching, but did not fully consider the influence of other people's actions, such as moving objects out of the monitored environment. Yagi et al. [23] developed a system based on a wearable camera to assist object searches by displaying the last scene of the detected target object. However, the system's functionality is undermined if someone else has moved the object. Prior to our developmental research, we explored the situation where people are unable to find something in a group of people. Questionnaire feedback showed that moving objects by oneself is a very common reason for not finding things. For example, if the image captured in the last frame of camera system is of you, you may not be able to find the object because you put it in a blind spot. However, objects being moved or carried away by others is also an important cause. If the image captured is of someone else, it may be because they moved the object and it is not in its original position.

For this reason, we designed and developed a prototype called CamFi to help us find things in public places. Our system consists of a camera, a server and a smartphone application. The camera is used to obtain video stream input. By using technology of object detection and face recognition, the server processes the frames of the video stream. The scene in which the object was last seen, who was last seen with the object and who was most often seen with the object can be recorded. Users can obtain all of this information by looking up the type of object on a smartphone and clicking on a thumbnail image.

The usability experiment was conducted in a room in our laboratory. In a simulated multi-person scenario, 12 people used the application. The purpose of our experiment is to prove that our system can effectively help users find lost objects in multi-person scenarios. At the end of the experiment, we asked participants to complete questionnaires and also conducted semi-structured interviews.

The contribution of the paper is:

- Through a questionnaire survey, we summarized the main reasons why users lose objects in public places
- We proposed a solution of assisted object finding based on a camera and artificial intelligence (AI) technology in a multi-person scene, and designed and developed the corresponding prototype.
- We conducted experiments and demonstrated the usability of our system in helping users find lost items in multi-person scenarios through quantitative and qualitative analysis.

2 RELATED WORK

This paper brings together related work on previous computational systems for finding lost objects and camera-based systems in public space.

The human-computer interaction (HCI) field has a long-standing research interest in harnessing sensor technologies, including wireless tags [11, 13, 20, 21], Bluetooth [10, 15, 19] and cameras [4, 22], to help users find lost objects. Previously, researchers have tended to deploy wireless tags, i.e., radio-frequency identification tags, on objects [13, 21] and allow users to locate a target item by moving through a given range. However, it is too time-consuming and unrealistic to attach an external tag to every object. To solve this problem, Yagi et al. developed a registration-free wearable system for helping users to find items they had held in their hands [23]. However, they assumed each object was manipulated by a single user and did not consider multi-user scenarios in public spaces where objects can be moved by other people.

Stationary cameras or closed-circuit television (CCTV) cameras are common in public spaces [14] and typically serve as security assurance [9]. In the HCI community, researchers often exploit the potential of public cameras to detect, track and interact with people in a public space [5, 6, 12]. For example, Prange et al. explored the potential of ubiquitous technologies like cameras to secure personal items in public spaces [18]. When we lose some items, we tend to look for them by reviewing the security camera footage, which is very painstaking and time-consuming. Nevertheless, cameras still offer potential for locating items. Xie et al. designed a dual-camera sensor network for object retrieval, but did not consider the case where targeted objects were obscured or taken away [22].

This previous research inspired us to develop a camera-based system to address the problem of finding lost items in public places. Therefore, our work develops the field by implementing AI algorithms to help users find lost items in multi-person scenarios without the need to pre-register or attach any other devices (i.e., only use stationary cameras) and to provide an intuitive retrieval method via their mobile phones.

3 FORMATIVE INVESTIGATION

Before starting our study, 125 questionnaires were given to people, asking for the most common reasons why people cannot find an object in a public place like an office. One or more answers were allowed. Of the returned questionnaires, five were invalid (such as blank), therefore we analyzed the remaining 120 (64 males and 56 females). The people who submitted valid questionnaires ranged in age from 18 to 60 (M= 31.47, SD=8.74). We identified 10 reasons through open coding: (1) the object has been left unused for too long; (2) the object does not catch people's attention; (3) the object was taken by me but forgotten; (4) I was interrupted while handling the object; (5) unconscious actions affect the position of objects; (6) I put objects in blind spots such as in a drawer; (7) objects have been moved by others; (8) objects are moved by non-human causes; (9) disease impacts memory (e.g., Alzheimer's, amnesia, etc.); (10) other reasons.

A frequency for each answer is shown in Figure 1. After classification, we found that the overall reasons can be classified into four categories: 43.2% are caused by the individual's own behavior; 30.6% are caused by long-term non-use or relate to unimportant objects; 17.9% are caused by other people's actions; 3.1% are caused by some other uncommon reasons.

We used this information to design and develop a search system for multi-person scenes.

4 SYSTEM DESIGN

4.1 Interaction

First, images need to be taken and names recorded of people who frequent the public scene. These photos and names can be displayed later when information about the person is needed.

The interface of the application is shown in Figure 2. When searching for objects, users can open the application and enter the type of object they want to search for in the input box. When the search command is executed, all thumbnails of that type of object appear below the search box. The user can then click on the thumbnail to jump to the details page and can see a complete picture of the object in its scene. At the bottom of the object image, we provide recommendations related to people, including the last person to be seen in the scene and the person who has been seen most often with the object. After clicking on the people image, the list of people is displayed. There may be more than one person in the final scene and there may be more than one person who has had the object most often, therefore, a list of possible options is displayed.

Through display box A, the last position of the object recorded by the camera can be quickly identified. If the object is not in the main display, there are two possibilities: (1) the object has been placed in a blind spot, such as behind another object or in a drawer. Since we record the photos at a relatively short intervals (2 second [s]), we can see the person's movements before placing the object and can infer the possible final position of the object; (2) the object has been taken away from the environment. In this case, we can find the last person to use the object by looking at display box B. Finally, if the lost object still cannot be found, display box C shows who is likely to use the object most often. This person may have a better idea of where objects are often placed or be able to provide other valuable suggestions.

If there is more than one recent user and more than one most frequent user, display boxes B and C will show those people with the highest priority. After clicking display box B, the names and pictures of all recent users will be displayed, ordered by their proximity

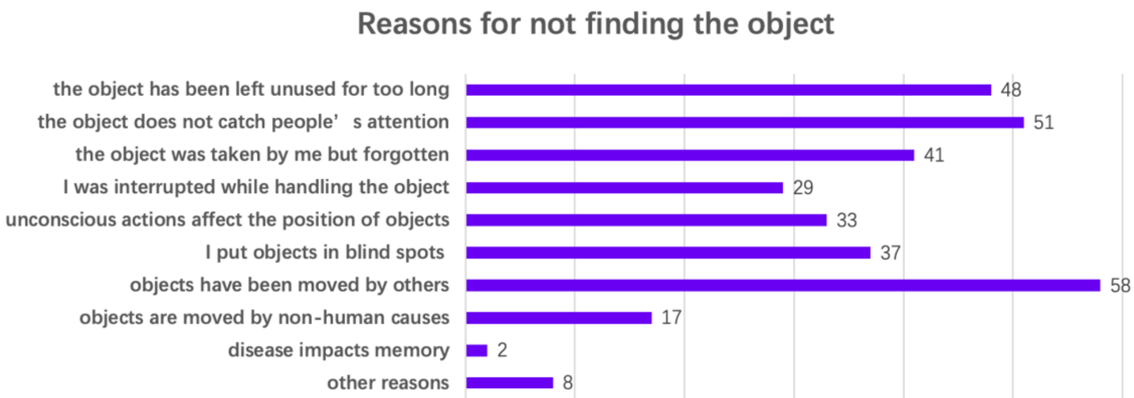

Figure 1: Reasons For Not Finding Objects in Public Places

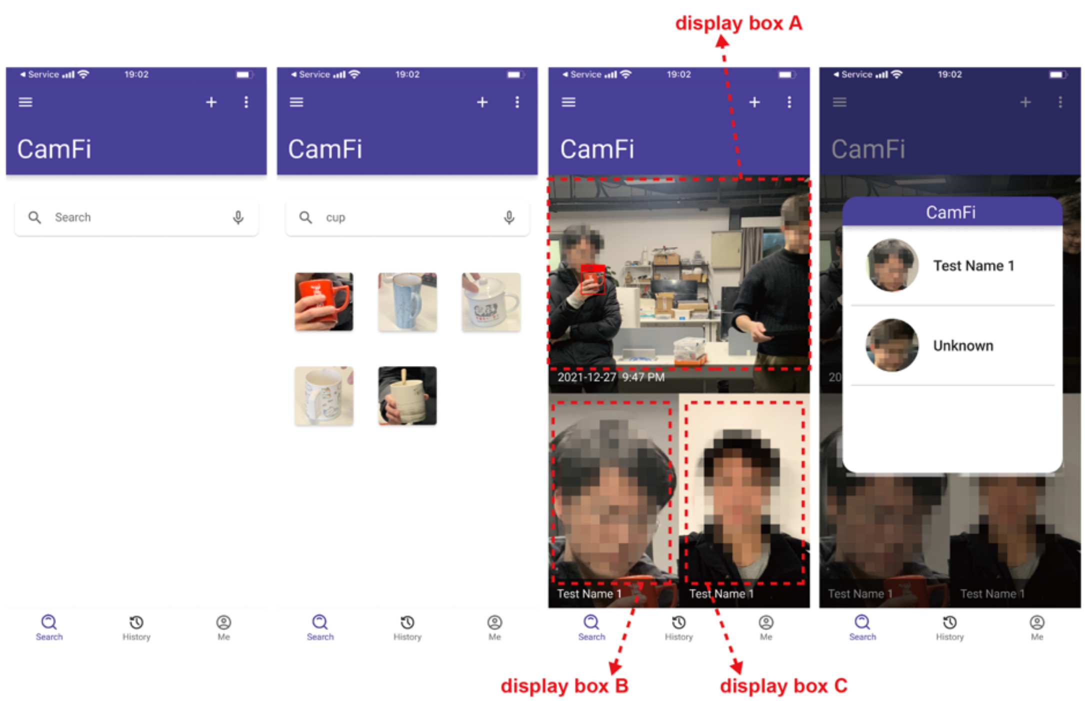

Figure 2: Application Interface (display box A shows the last scene image, display box B shows the last user image, display box C shows the most common user image)

to the object according to the distance between the coordinates of their face and the coordinates of the object. The names of permanent office staff will be displayed and strangers will be tagged as 'unknown'. After clicking display box C, the names and images of all those who most frequently used the object are displayed, in the order of most recent according to the last time they were seen with the object. The list in display box C shows only known office staff.

4.2 Implementation

4.2.1 System Overview. CamFi requires a camera, processing server and smartphone for browsing the location of objects. The camera type is OAK-D-POE and the MySQL database has been installed on the server. The camera feeds a live video stream over a local area network (LAN) to the server, which processes each frame every 2s and saves the necessary information. The user requests access to the server through the smartphone and the server will return the thumbnail, panorama and relevant person information to the user. This system has universality and other cameras can achieve the

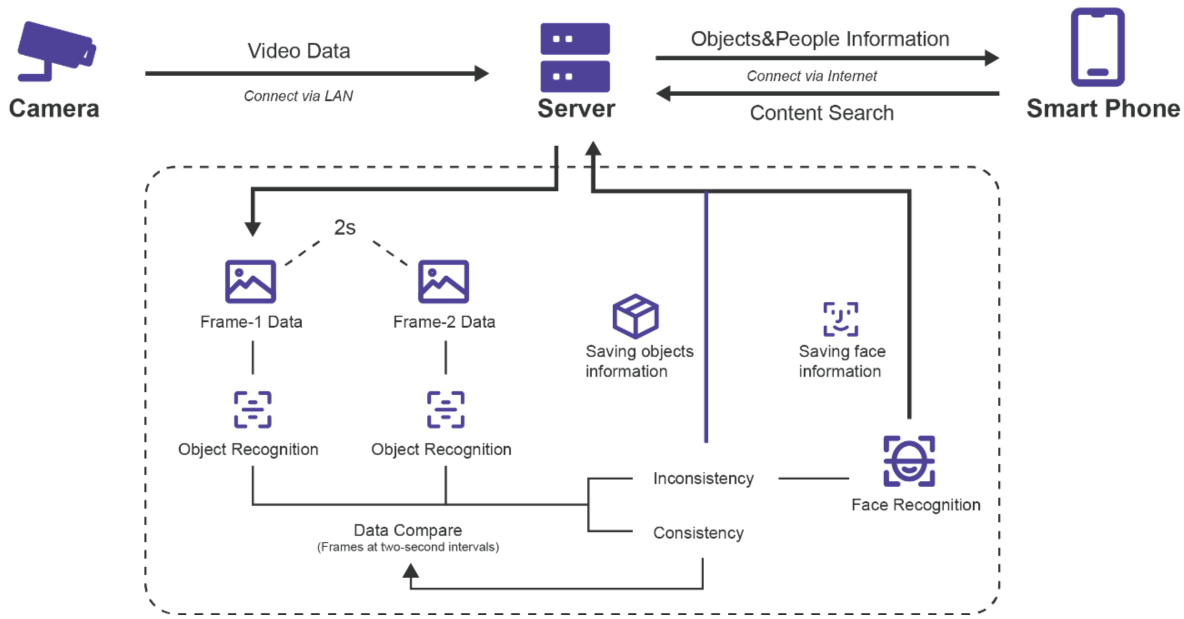

Figure 3: System Diagram

same function with the algorithm used in the system. The whole system is shown in Figure 3

4.2.2 Object Detection. ImageNet [7] has been used as the image dataset. We select YOLOv4[1] [2] as the object detector to process the real-time video stream obtained by the camera. Instead of processing every frame in the video stream, we take a frame every 2s and carry out multi-object recognition detection on the image of this frame. Rectangular box labeling is carried out on the recognized object and coordinates are obtained. Each frame is compared with the frames taken 2s before and after. If there is no newly identified object and the coordinates of the object have not changed, the position of the object has not changed and the new frame will not be saved. Otherwise, the picture is saved. Then, the image is sent to a face recognition algorithm for processing to obtain all the face information in the image. If it is a registered member of staff, the face recognition algorithm can obtain their name, otherwise it will be marked as unknown.

Finally, the changing information in the picture is recorded in the MySQL database, including the type and coordinates of the object, and the name of staff member. For every new or moving object, a piece of data is generated.

4.2.3 Face Recognition. We used models from OpenCV for face recognition. First, the permanent office staff will need to upload a close-up, full-frontal photo to the face library. The specific identification process is as follows. First, the face-detection-retail-0004 model was run to detect the faces in the images and intercept the face images. Then, the landmarks-regression-retail-0009 model was run to detect facial feature points. Following this, the face-reidentification-retail-0095 model was run to obtain the feature vectors and the obtained image feature vectors were added to the list. Finally, the cosine distance was calculated between the feature vector of the picture in the face library and the feature vector of the image obtained from the camera. The closer the cosine distance, the higher the similarity.

When the user searches, the picture in which the object appeared for the last time can be obtained from the database and is shown in display box A. After this picture is processed by the face recognition algorithm, the faces of all the people in this picture can be obtained, including permanent staff and strangers, and the highest priority people will be shown in display box B. At the same time, according to the information stored in the database, the staff members who are most often seen with objects are also identified and those of the highest priority will be shown in the display box C.

5 USER STUDY

5.1 Participants and Procedure

We recruited 12 volunteers (7 males and 5 females) with ages ranging from 23 to 40 (M=28.5, SD=4.94). The study was performed in our laboratory.

First, all participants uploaded full-frontal photos of themselves. Then participants were given a five-minute introduction to the system's functions and general structure, followed by a five-minute demonstration of how the system works. They were asked to place a total of 30 common laboratory objects, such as umbrellas, notebooks and cups, in different positions. They were then asked to spend 30 minutes walking in and out of other rooms and doing 20 simple calculations in order to interfere with their memory of where objects had been placed. The 20 questions were mathematical calculations involving four numbers ranging from 1 to 10, including addition, subtraction, multiplication and division. If less than 90 percent were correct, participants were asked to complete 20 new

[1]https://pjreddie.com/darknet/yolo/

Figure 4: Participants Interacting with CamFi

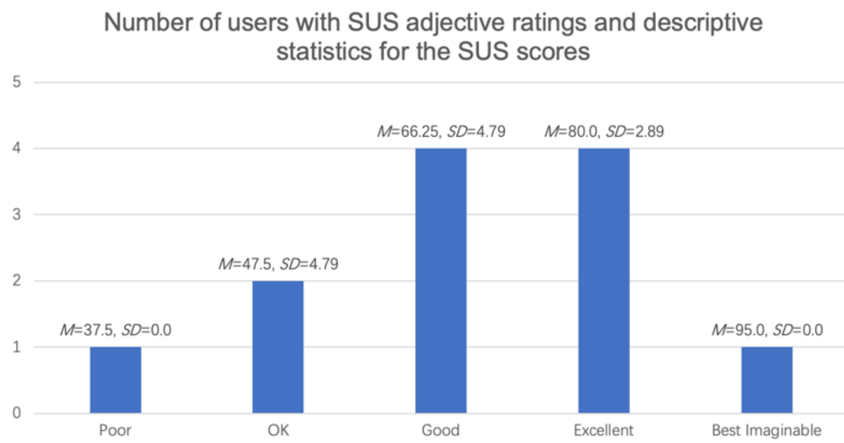

Figure 5: System Usability Scale Results

calculations to avoid scribbling their answers. During the interval, other researchers randomly removed some of the objects to simulate a real public scene. Finally, the participants were able to use the system to find the objects they had placed or the researchers who had removed them (see Figure 4). In the course of this experiment, face recognition was all correct.

After the experiment, we asked the participants to fill in the SUS (System Usability Scale) questionnaire [1, 3]. We then conducted semi-structured interviews around the following five questions: Q1. What do you think is the difference between this system and traditional surveillance cameras? Q2. Do you think the system prompts the most frequent or most recent person to help you? Q3. Do you expect the system to be used on a large scale? Q4. Which scenarios do you think are suitable for using this system? Q5. Where do you think the system needs to be improved?

5.2 Quantitative Results

The system is supposed to be easy to learn. According to the SUS questionnaire feedback results, most people found the system easy to use and did not need additional technical support. As shown in Figure 5, nine participants rated its usability as good or better overall. One of them gave a best imaginable rating. On the whole, the feedback results confirmed its usability.

5.3 Qualitative Results

In the semi-structured interviews, participants generally believed that the system was very useful. Compared with traditional surveillance cameras, most people thought this system is convenient, saves time and has targeted functions. According to P5, it is necessary for the system to prompt information of the most frequent and recent people in the scene involving multiple people, while P4 stated out that the main scene could already meet most of the requirements. If this system were to be widely used, P8 felt it would be good, but P6 and P7 felt it may violate privacy and hoped that privacy issues will be considered. In terms of the use scenario, P1 stated that it may be suitable for locations such as a chemical laboratory and a storeroom. Participants also provided many suggestions to the researchers, among which were suggestions from P3 and P5 that multiple cameras could be added to reduce blind spots.

Overall, supplementing face information was considered a useful strategy for finding lost objects in public settings. Privacy concerns raised by participants were also considered in advance, and we planned to limit this system to certain public settings where privacy were not high.

6 LIMITATIONS AND FUTURE WORK

There are still some areas that can be improved in the system. In the order of the list of recent users, we set the priority order based on the distance between the coordinates of the face and the object in the picture. However, the camera angle may be different from the real-world situation and different heights will also affect the priority order result. In addition, this system inevitably involves privacy issues. However, we plan to use this system only in public places, especially those where privacy is not high, such as storage rooms and chemistry labs.

In this study, we only conducted a preliminary experiment on the effectiveness of the system. In the future, we plan to set up control groups to carry out further experiments, including a three-arm design with an experimental group using the system, a control group using other finding functions and a control group without any tools. We will conduct statistical analysis of the time and success rate in finding targets in multi-user scenarios. In addition, we plan to optimize the face recognition algorithm. Under the current system, the information regarding the faces of staff members needs to be pre-recorded under the condition of close distance without occlusion. After optimization, when the distance and other requirements are not so high, this step could be performed automatically when the camera detects a face, making the operation of the whole system more concise and smoother.

7 CONCLUSIONS

This article provides a solution for finding objects in a multi-person scene. The system consists of a camera, a server and a mobile application. Object recognition and face recognition technology can help users search for objects in multi-person scenes and provide information about recent users and frequent users. In this way, when objects are still not found via the main scene, people can find them more easily by finding the most recent or most frequent users. In summary, we designed and developed CamFi and conducted a preliminary usability study. Quantitative analysis and qualitative analysis prove the usability of the system. CamFi may be an effective tool to help users find items in a multi-user scenario.

ACKNOWLEDGMENTS

This research was funded by the Engineering Research Center of Computer Aided Product Innovation Design, Ministry of Education, Fundamental Research Funds for the Central Universities, National Natural Science Foundation of China (Grant No. 52075478), and National Social Science Foundation of China (Grant No. 21AZD056).

REFERENCES

[1] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual sus scores mean: adding an adjective rating scale. J. Usability Stud. 4, 3 (May 2009), 114–123.

[2] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao. 2020. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv: 2004.10934. Retrieved from https://doi.org/10.48550/arXiv.2004.10934

[3] John Brooke. 1996. SUS: a "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), Usability Evaluation in Industry. Taylor and Francis.

[4] Andreas Butz, Michael Schneider, and Mira Spassova. 2004. SearchLight – a lightweight search function for pervasive environments. In Pervasive Computing (Lecture Notes in Computer Science). Springer, Berlin, Heidelberg, 351–356. https://doi.org/10.1007/978-3-540-24646-6_26

[5] Sunsern Cheamanunkul, Evan Ettinger, Matt Jacobsen, Patrick Lai, and Yoav Freund. 2009. Detecting, tracking and interacting with people in a public space. In Proceedings of the 2009 international conference on Multimodal interfaces (ICMI-MLMI '09). Association for Computing Machinery, New York, NY, USA, 79–86. https://doi.org/10.1145/1647314.1647330

[6] Anooshmita Das, Krister Jens, and Mikkel Baun Kjærgaard. 2020. Space utilization and activity recognition using 3d stereo vision camera inside an educational building. In Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers (UbiComp-ISWC '20). Association for Computing Machinery, New York, NY, USA, 629–637. https://doi.org/10.1145/3410530.3414318

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: a large-scale hierarchical image database. In 2009 IEEE Conference on Computer Vision and Pattern Recognition. 248–255. https://doi.org/10.1109/CVPR.2009.5206848

[8] Margery Eldridge, Abigail Sellen, and Debra Bekerian. 1992. Memory Problems at Work: Their Range, Frequency and Severity. Rank Xerox EUROPARC. Retrieved from https://europe.naverlabs.com/research/publications/memory-problems-at-work-their-range-frequency-and-severity/

[9] Nicholas R. Fyfe and Jon Bannister. 1996. City watching: closed circuit television surveillance in public spaces. Area 28, 1 (1996), 37–46.

[10] Julie A. Kientz, Shwetak N. Patel, Arwa Z. Tyebkhan, Brian Gane, Jennifer Wiley, and Gregory D. Abowd. 2006. Where's my stuff? design and evaluation of a mobile system for locating lost items for the visually impaired. In Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility (Assets '06). Association for Computing Machinery, New York, NY, USA, 103–110. https://doi.org/10.1145/1168987.1169006

[11] Mizuho Komatsuzaki, Koji Tsukada, Itiro Siio, Pertti Verronen, Mika Luimula, and Sakari Pieskä. 2011. IteMinder: finding items in a room using passive rfid tags and an autonomous robot (poster). In Proceedings of the 13th international conference on Ubiquitous computing - UbiComp '11. ACM Press, Beijing, China, 599. https://doi.org/10.1145/2030112.2030232

[12] Sandjar Kozubaev and Carl DiSalvo. 2021. Cracking public space open: design for public librarians. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, 1–14. Retrieved from https://doi.org/10.1145/3411764.3445730

[13] Xiaotao Liu, Mark D. Corner, and Prashant Shenoy. 2006. Ferret: rfid localization for pervasive multimedia. In Proceedings of the 8th international conference on Ubiquitous Computing (UbiComp'06). Springer-Verlag, Berlin, Heidelberg, 422–440. https://doi.org/10.1007/11853565_25

[14] Clive Norris and Jade Moran. 2016. Surveillance, Closed Circuit Television and Social Control. Routledge.

[15] Ling Pei, Ruizhi Chen, Jingbin Liu, Tomi Tenhunen, Heidi Kuusniemi, and Yuwei Chen. 2010. Inquiry-based bluetooth indoor positioning via rssi probability distributions. In Proceedings of the 2010 Second International Conference on Advances in Satellite and Space Communications (SPACOMM '10). IEEE Computer Society, USA, 151–156. https://doi.org/10.1109/SPACOMM.2010.18

[16] Rodney E. Peters, Richard Pak, Gregory D. Abowd, Arthur D. Fisk, and Wendy A. Rogers. 2004. Finding Lost Objects: Informing the Design of Ubiquitous Computing Services for the Home. Georgia Institute of Technology. Retrieved from https://smartech.gatech.edu/handle/1853/51

[17] Pixie Technology. 2017. The nation's biggest lost and found survey by pixie. Retrieved January 12, 2022 from https://tinyurl.com/yxrzbsnp

[18] Sarah Prange, Lukas Mecke, Michael Stadler, Maximilian Balluff, Mohamed Khamis, and Florian Alt. 2019. Securing personal items in public space: stories of attacks and threats. In Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia (MUM '19). Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3365610.3365628

[19] David Schwarz, Max Schwarz, Jörg Stückler, and Sven Behnke. 2015. Cosero, find my keys! object localization and retrieval using bluetooth low energy tags. In RoboCup 2014: Robot World Cup XVIII (Lecture Notes in Computer Science). Springer International Publishing, Cham, 195–206. https://doi.org/10.1007/978-3-319-18615-3_16

[20] Masaya Tanbo, Ryoma Nojiri, Yuusuke Kawakita, and Haruhisa Ichikawa. 2017. Active rfid attached object clustering method with new evaluation criterion for finding lost objects. Mob. Inf. Syst. 2017, (February 2017), e3637814. https://doi.org/10.1155/2017/3637814

[21] Paul Wilson, Daniel Prashanth, and Hamid Aghajan. 2007. Utilizing rfid signaling scheme for localization of stationary objects and speed estimation of mobile objects. In 2007 IEEE International Conference on RFID. 94–99. https://doi.org/10.1109/RFID.2007.346155

[22] Dan Xie, Tingxin Yan, Deepak Ganesan, and Allen Hanson. 2008. Design and implementation of a dual-camera wireless sensor network for object retrieval. In Proceedings of the 7th international conference on Information processing

in sensor networks (IPSN '08). IEEE Computer Society, USA, 469–480. https://doi.org/10.1109/IPSN.2008.57

[23] Takuma Yagi, Takumi Nishiyasu, Kunimasa Kawasaki, Moe Matsuki, and Yoichi Sato. 2021. 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. ACM, College Station TX USA, 139–149. https://doi.org/10.1145/3397481.3450664