Anotoki Camera: A Memory Support System Based on Egocentric Image Archiving and Retrieval

1st Keiko Kodama
Graduate School of Engineering
Kobe University
1-1 Rokkodai-cho, Nada, Kobe, 657-8501, Japan
kodakoma@es4.eedept.kobe-u.ac.jp

2nd Ryota MURATE
Graduate School of Engineering
Kobe University
1-1 Rokkodai-cho, Nada, Kobe, 657-8501, Japan
kodakoma@es4.eedept.kobe-u.ac.jp

3rd Shun HIRAI
Graduate School of Engineering
Kobe University
1-1 Rokkodai-cho, Nada, Kobe, 657-8501, Japan
kodakoma@es4.eedept.kobe-u.ac.jp

4th Takuya NAKATA
Center of Mathematical and Data Sciences
Kobe University
1-1 Rokkodai-cho, Nada, Kobe, 657-8501, Japan
chensinan@gold.kobe-u.ac.jp

5th Sinan Chen
Center of Mathematical and Data Sciences
Kobe University
1-1 Rokkodai-cho, Nada, Kobe, 657-8501, Japan
chensinan@gold.kobe-u.ac.jp

6th Sachio Saiki
School of Data and Innovation
Kochi University of Technology
185 Miyanokuchi, Tosayamada, Kami, Kochi, 782-8502, Japan
saiki.sachio@kochi-tech.ac.jp

7th Masahide Nakamura
Center of Mathematical and Data Science
Kobe University
1-1 Rokkodai-cho, Nada, Kobe, 657-8501, Japan
masa-n@cmds.kobe-u.ac.jp

8th Kiyoshi Yasuda
Graduate School of Engineering
Riken AIP
1-4-1 Nihon-bashi, Chuo-ku, Tokyo 103-0027 Japan
yasukiyo.12@outlook.jp

***Abstract*—In Japan, the prevalence of mild cognitive impairment (MCI) is rising with population aging, and episodic-memory deficits burden individuals and caregivers. Various memory support methods have been proposed that collect and present daily life data, but many rely on third-person perspective data or require complex operations and prior knowledge. Given this background, there is a growing need for systems that enable individuals with MCI to independently collect, store, and retrieve data that facilitates episodic memory recall. Previous studies have proposed systems that capture and store egocentric images and enable tag-based search, but some issues remain. To address these challenges, this study proposes a new memory support system called Anotoki Camera, which enhances a previous prototype by automatically filtering images and utilizing LLM for tagging. We implemented the system and conducted a case study, confirming that it can automatically eliminate unnecessary images and assign tags useful for search.**

***Index Terms*—episodic memory, aging population, memory impairment, mci, lifelog**

I. Introduction

In Japan, the aging population is rapidly increasing, with the proportion of individuals aged 65 and older reaching 29.1% of the total population in 2023, the highest level in the world [1]. Consequently, the prevalence of dementia and mild cognitive impairment (MCI) is also increasing annually, posing a serious challenge not only for individuals and families but also for society as a whole [2]. MCI is a condition that precedes dementia, where individuals can mostly live independently but experience mild impairments in memory and attention [3]. In particular, recent episodic memory is often impaired, leading to everyday problems such as ”forgetting whether they took their medication” or ”not remembering where they placed items.” Episodic memory refers to the memory of events that an individual has personally experienced, and since its content is unique to the individual, it is difficult for even close family members to compensate for it. Therefore, not only do individuals with memory impairments experience anxiety and frustration, but those around them also face psychological and time burdens, such as repeatedly answering the same questions and dealing with unexpected behaviors [4]. Based on this background, there is a growing need for systems that enable elderly individuals with MCI to independently collect, store, and retrieve data that facilitates episodic memory recall. Especially, first-person visual information that closely resembles the user’s own perspective has been reported to be effective as a trigger for memory recall [5], highlighting the need for natural image acquisition and utilization in daily life.

To address these challenges, this study proposes a system called ”Anotoki Camera” that automatically captures and stores egocentric images and supports memory recall by allowing users to search and present images as needed. The system aims to enable elderly individuals to independently collect, store, and retrieve egocentric image data. Previous studies have proposed systems that capture and store first-person visual snapshots and enable tag-based search, but some issues remain.

- **P1:** Storage capacity is strained due to the high frequency of capturing and storing egocentric images.
- **P2:** Inadequate search capabilities due to low accuracy in automatic tagging.

Therefore, it is necessary to implement a system that avoids storing unnecessary images for episodic memory recall and applies sufficient tagging methods for effective search. To address these challenges, this study constructs a memory support system based on the following four approaches.

- A1: Exclude unclear images using rectangle detection.
- A2: Exclude similar images using inter-image similarity scores.
- A3: Tag Generation Using LLM.
- A4: Provide a web service for tag-based image search.

This paper implements a system based on the above approaches. In a case study, we confirmed that the implemented system can collect, store, and retrieve egocentric images. We also confirmed that the exclusion of similar images based on similarity scores and the exclusion of unclear images using rectangle detection are performed automatically. Furthermore, we confirmed that the tag generation using LLM assigns tags useful for search.

II. Background

A. MCI and Its Current Status

In Japan, the aging population is progressing, with the proportion of individuals aged 65 and older reaching 29.1% of the total population in 2023, the highest level in the world, and this trend is expected to continue [1]. In this super-aging society, the prevalence of MCI is increasing annually, becoming a social issue [2]. MCI is a condition that precedes dementia, where basic daily activities are normal, but there are complaints of cognitive decline and memory impairment from the individual or family [3]. In MCI, recent episodic memory is particularly impaired, leading to forgetfulness about actions such as whether they took their medication or had a meal, or where they placed items. Episodic memory refers to the memory of events that an individual has personally experienced, and it cannot be easily compensated for by close family members. Therefore, memory impairments in MCI can lead to anxiety and frustration, as well as difficulties in daily life. Additionally, the burden of repeatedly answering the same questions and dealing with unexpected behaviors makes care by family members and caregivers challenging [4]. Therefore, to address MCI, there is a need for systems that can automatically collect and present life records from the perspective of the elderly, allowing them to independently supplement gaps in their episodic memory.

B. Episodic Memory and Visual Data

Episodic memory is the memory of events and scenes that an individual has personally experienced, and the recall of such memories often relies on situational and visual information as important cues. Especially for elderly individuals with mild cognitive impairment (MCI), it has been reported that visual information naturally obtained in daily life is effective in supporting memory. Among these, first-person perspective image data that closely resembles the user’s own viewpoint is considered to have high acceptability for the individual [5].

In this paper, images close to the user’s first-person perspective are referred to as ”egocentric images” and the goal is to collect and utilize these images in daily life. In recent years, lifelogging technologies that collect various data such as images, videos, audio, and accelerometer sensors have been developed. However, in studies targeting elderly individuals, the focus is often on ”monitoring” using fixed cameras or wearable sensors, and records from a subjective perspective are limited [6]. Also, methods for obtaining egocentric images, such as capturing images when grasping specific objects or using pre-registered target objects as triggers, have been proposed [7]. These methods lack flexibility and are difficult for elderly individuals who are unfamiliar with operations to use independently. Therefore, there is a need for technology that can continuously capture and utilize images close to the user’s natural field of view without special operations.

C. Previous Studies

1) Overall Research Objective: The ultimate goal of this research is to support elderly individuals with MCI in independently supplementing their episodic memory. There are three main issues to address in achieving this goal.

- G1: Elderly individuals are not capturing egocentric image data.
- G2: Elderly individuals are unable to independently collect and store data.
- G3: Elderly individuals are unable to independently retrieve data that facilitates episodic memory recall.

To address these three issues, this research aims to develop a system that enables elderly individuals with MCI to independently collect, store, and retrieve egocentric image data that facilitates episodic memory recall.

2) Overall System Architecture Aimed by This Research: oVer all system architecture of this research is shown in Figure 1. Users wear sensing devices during their daily lives, and egocentric images are automatically captured during this time. The captured images are automatically tagged with metadata such as the time and location of capture, as well as tags related to objects and situations depicted in the images. By combining this information and storing it in a searchable database, an episode pool is formed to facilitate episodic memory recall. Users can search and present information from this database as needed, providing triggers to help them recall past events. By constructing such a system, we aim to address the three issues mentioned in II-C1 and provide memory support for elderly individuals with MCI.

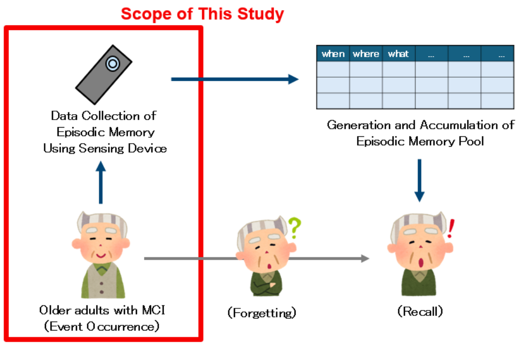

Fig. 1: Overall System Architecture Aimed by This Research

D. Prototype System Using PhotoPrism

In previous studies, a prototype system was proposed that focused on the area enclosed by the red frame in Figure 1, specifically aiming to address issues G1 and G2. The system uses a wearable device consisting of a Raspberry Pi Zero 2 W and a camera module to automatically capture egocentric images approximately every 3 seconds and save them to online storage via Wi-Fi. Furthermore, PhotoPrism [9] is used to automatically tag objects and situations depicted in the images, enabling search based on tags and temporal information. Users can access the online storage through a web browser to search and view the saved images. The wearable device was actually built, and Nextcloud [8] was used as the storage, while PhotoPrism [9] was used for tagging and search functionality. In the case study, it was confirmed that egocentric images could be automatically captured at regular intervals, tagged with information at the time of capture, and saved to Nextcloud. On the other hand, issues related to storage capacity and tagging accuracy became apparent.

E. Issues in Previous Studies

In the prototype system proposed in previous studies, the following two issues were identified. **Issue P1: Storage capacity is strained due to the high frequency of capturing and storing egocentric images Issue P2: Inadequate search capabilities due to low accuracy in automatic tagging**

F. Scope of This Research

This research aims to develop a system that enables elderly individuals with MCI to independently collect, store, and retrieve egocentric data that facilitates episodic memory recall, based on the objectives and scope described in II-C1. This paper focuses on issues P1 and P2 identified in previous studies. The challenges to be addressed from these issues include implementing a mechanism that avoids storing unnecessary images for episodic memory recall and applying sufficient tagging methods for effective search.

III. Proposed Method

A. Objectives and Approaches

The objective of this research is to address the issues identified in the prototype system proposed in previous studies and to develop an effective system for memory support for elderly individuals with MCI. To achieve this, this paper proposes a memory support system called ”Anotoki Camera.” The key idea is to build a new system based on the existing prototype system by incorporating automatic image selection and large language models (LLM). While the system is based on the prototype system, it aims to resolve issues P1 and P2 by introducing the following four new approaches.

- A1: Exclude unclear images using rectangle detection.
- A2: Exclude similar images using inter-image similarity scores.
- A3: Tag Generation Using LLM.
- A4: Provide a web service for tag-based image search.

B. Overall Architecture

The overall architecture of the proposed system is divided into two phases: the saving phase and the searching phase. The architecture of the saving phase is shown in Figure 2, and the architecture of the searching phase is shown in Figure 3. The proposed system consists of a wearable device with a camera module, a server, and a database. On the server, API integration with LLM for tag generation for captured images and a web service that enables users to search images are operational, facilitating image saving and searching for memory support through these components. Users wear the wearable device around their necks and go about their daily lives, during which egocentric images are automatically captured, processed, and saved. The following sections describe the flow of processing in each phase.

*1) **Saving Phase**:*

***Step 1: Capture egocentric snapshots with a Wearable Device**:* Users wear a wearable device with a camera module around their necks and go about their daily lives. By wearing it around the neck, images from a first-person perspective that closely resemble the user's actual field of view can be recorded. The wearable device automatically captures egocentric images at regular intervals after being powered on. This allows users to continuously obtain egocentric images without being conscious of special operations, enabling even elderly individuals who are unfamiliar with operating machines to use the system without burden.

***Step 2: Adding Capture Time Information**:* The images obtained in Step 1 are automatically tagged with the date and time of capture. This is to enable the use of temporal cues, such as ”when it was captured,” in addition to visual information for searching. In episodic memory recall, there are many situations where temporal context is important in addition to visual information, making it useful for enhancing the usability of the data.

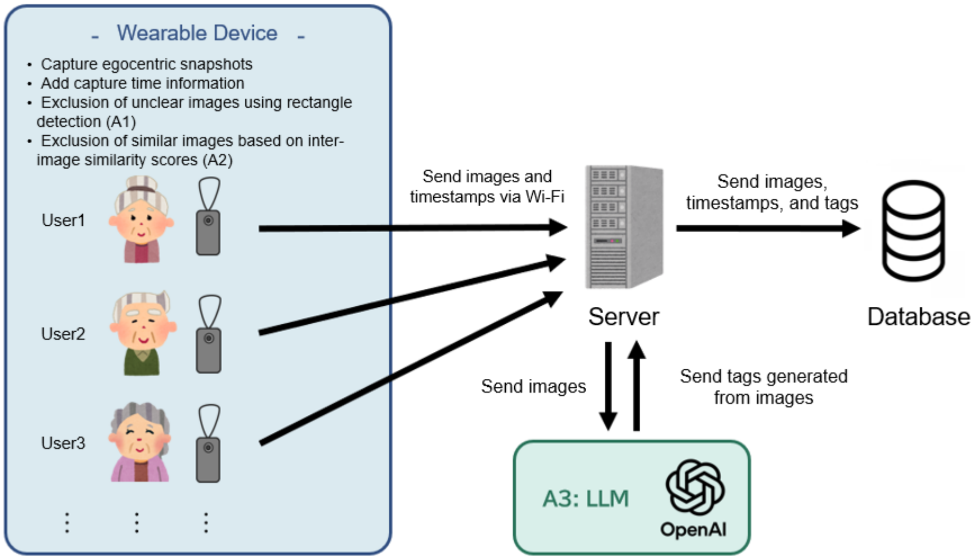

Fig. 2: Anotoki Camera: Architecture of the Saving Phase(A1, A2, A3)

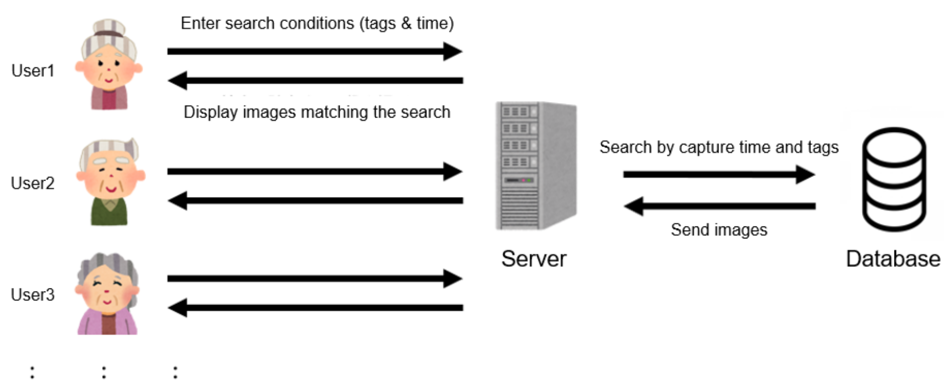

Fig. 3: Anotoki Camera: Architecture of the Searching Phase(A4)

Step 3: Exclude unclear images using rectangle detection (A1): Unclear images among the captured images are automatically excluded based on rectangle detection. Since the device continues to capture images at regular intervals while the user is walking or performing daily activities, a certain number of unclear images that are difficult to recognize due to body movement-induced blur or focus misalignment are captured. Such unclear images are relatively more likely to be assigned incorrect tags that do not match the content of the image and are less likely to facilitate episodic memory recall. Additionally, continuously accumulating unnecessary images for support leads to the strain of limited storage capacity. In this system, the number of rectangles in the image is extracted using rectangle detection, and images with an extremely low or excessively high number of rectangles are determined to be unclear images, and subsequent processing is not performed on them.

Step 4: Exclude similar images using inter-image similarity scores (A2): Similar images among the captured images are automatically excluded based on inter-image similarity scores. Since the device is designed to automatically capture the field of view at regular intervals, during periods when the user is stationary or in situations with minimal movement, images with similar compositions and content may be continuously saved. Such similar images are redundant in supporting memory recall and lead to the accumulation of a large number of low-information-value data. In particular, when a user attempts to search for past events, if a large number of images with almost identical content are displayed as search results, it may become difficult to find the necessary images and hinder memory recall. Additionally, continuously saving a large number of semantically redundant images leads to the strain of storage capacity. In this system, inter-image similarity scores are calculated, and images that are determined to be highly similar are automatically excluded. Specifically, the similarity of consecutive images is calculated based on Structural SIMilarity (SSIM), and highly similar images that exceed a certain threshold are considered redundant and auto-

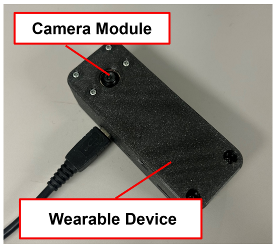

Fig. 4: Appearance of the Device

matically excluded.

Step 5: Tag Generation Using LLM (A3): After processing Steps 1 to 4 on the wearable device, the images are automatically uploaded to the server via Wi-Fi. In this step, tags representing objects depicted in the uploaded images are automatically generated using large language models (LLM). This enables users to accumulate egocentric images in a searchable format without the need for operations such as using a PC. By utilizing LLM, it becomes possible to automatically generate abstract and contextual labels. For example, in addition to object-level tags such as "chair," "desk," "computer," and "person," contextual tags like "office" and "in a meeting" can also be assigned. The generated tags, along with the images and capture times, are stored in the database.

*2) **Searching Phase**:*

Step 1: Input of Search Conditions by Users: Users can search for relevant images based on tags and capture dates through the web service. The search conditions are sent to the server via the network, and a search process is executed on the database of the target user.

Step 2: Retrieval and Presentation of Search Results: On the server side, the database is searched based on the specified conditions, and images that match the criteria are extracted. During this process, tags generated by LLM in the saving phase and metadata associated with the images (e.g., capture time) are utilized. The extracted images are presented on the web service, allowing users to visually confirm them. This enables users to efficiently access images that aid in memory recall from a vast collection of images.

IV. Implementation

A. Technologies Used

The appearance of the implemented device is shown in Figure 4. As a wearable device, the Raspberry Pi Zero 2 W [10] designed by the Raspberry Pi Foundation was used. The Raspberry Pi Camera Module 3 (wide-angle) [11] was used as the camera module. In the prototype system, a standard field-of-view camera was used, but there was an issue of limited visibility. In this research, a wide-angle camera was used to improve the accuracy of visual information acquisition, allowing for a broader field of view to be recorded at once.

The following technologies were used to build the web service. Additionally, OpenAI's ChatGPT-4o was used as the LLM.

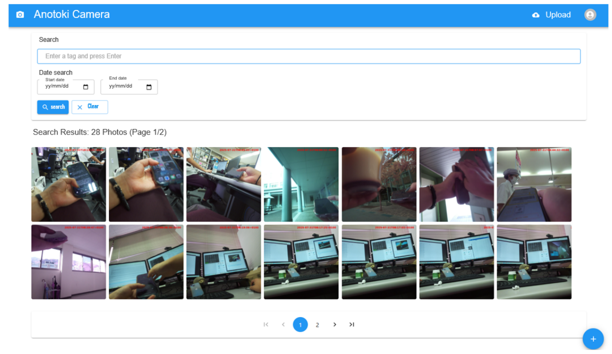

Fig. 5: Example of the Web Service Screen of Anotoki Camera

Fig. 6: Example of an Image Determined to be Unclear

- Development Languages: Python, JavaScript
- Storage: Nextcloud [8]
- Database: MySQL [12]

Nextcloud [8] is an open-source online storage platform that allows for file storage and sharing. It features high security and scalability, enabling flexible management of image data. The proposed system uses it for managing egocentric images.

The example of the web service screen that was actually created is shown in Figure 5. On this service, image searches can be performed based on specified tags and dates, and relevant images are displayed in a list.

V. Case Study

As a case study, the flow from saving to searching egocentric images was confirmed by actually using the Anotoki Camera in daily life.

A. Exclude unclear images using rectangle detection.

An example of an image that was determined to be unclear through rectangle detection among the actually captured images is shown in Figure 6. The image was determined to be unclear, so the tagging process was not performed.

B. Exclude similar images using inter-image similarity scores.

An example of two images that were actually captured and the similarity score calculated for them is shown in Figure 7. The score was high at 82.7%, and one of the images was determined to be unnecessary, so the tagging process was not performed.

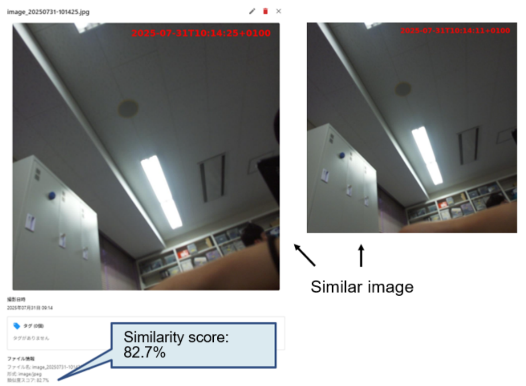

Fig. 7: Example of Two Images Determined to be Similar

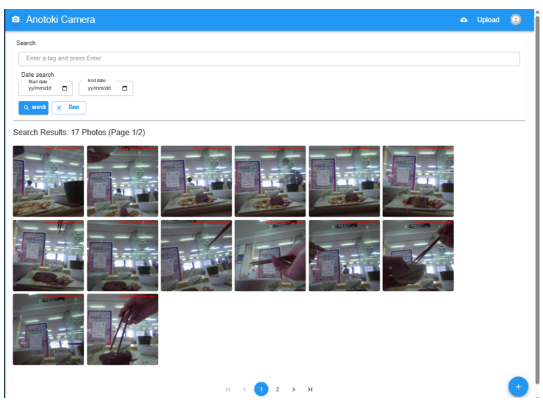

Fig. 8: Search Results Based on the Tag "Cafeteria" and Date

C. Tag Generation Using LLM

For each captured egocentric image, an LLM automatically generated about 20 diverse tags. For example, one image was assigned tags such as "lunch," "cafeteria," "Japanese food," "QR code," and "quiet atmosphere." These tags, reflecting objects, context, and atmosphere, are later used as cues for memory recall in the searching phase.

D. Image Search Based on Tags and Dates

Users can search for relevant images by specifying arbitrary tags and capture dates on the web service for images tagged by LLM. As an example, Figure 8 shows the screen of an actual search conducted on July 31, 2025, using the tag "cafeteria." As a result, egocentric images taken while having a meal in a university cafeteria were presented. All the presented images appropriately corresponded to the specified tags, suggesting that tag generation using LLM has a certain level of effectiveness. Additionally, it was confirmed that by utilizing both tags and dates for searching, users could quickly access their desired images from a total of 635 images.

VI. Discussion

Through the case study, several issues became apparent. First, while the tag search operation itself was performed without issues, search accuracy decreased in cases where synonyms or variations in notation were used. Therefore, flexible input processing, such as partial match searches, is considered necessary. Additionally, due to the mounting position of the camera used, there is inevitably a discrepancy between the actual line of sight and the recorded images. For instance, while the "cafeteria" scene was appropriately recorded, the meal contents were not sufficiently captured, making it difficult to confirm the details in the plate. To acquire information closer to the line of sight, it is necessary to consider using eyeglass-type devices or optimizing the mounting position.

VII. Conclusion

In this study, we proposed the "Anotoki Camera," a memory support system that implements a mechanism to avoid saving unnecessary images for searching episodic memory recall and applies an adequate tagging method for searching.

As future prospects, addressing notation variations, devising mounting methods to reduce discrepancies with the line of sight, and verifying in real-life scenarios that include objects frequently searched for by elderly individuals are mentioned.

Acknowledgment

This research was partially supported by JSPS KAKENHI Grant Numbers JP25H01167, JP25K02946, JP25K24389, JP24K02765, JP24K02774, JP23K17006, JP23K28091, JP23K28383.

References

[1] Cabinet Office, Government of Japan: Annual Report on the Ageing Society: (2024).
[2] Kyushu University: Study on the Prevalence and Future Estimation of Dementia and Mild Cognitive Impairment in Japan, pp. 9–12 (2024).
[3] R C Petersen and J C Morris: Mild Cognitive Impairment as a Clinical Entity and Treatment Target, Archives of Neurology, vol. 62, pp. 1160–1163 (2005).
[4] T Maruo, Y Ito, N Katakura, M Uda, N Ogawara: Everyday Challenges of People Diagnosed with Mild Cognitive Impairment and their Family Caregivers, Kobe City College of Nursing, vol. 26, pp. 57–65 (2022).
[5] O Gelonch, M Ribera, N Codern-Bové, S Ramos, M Quintana, G Chico, N Cerulla, P Lafarga, P Radeva, M Garolera: Acceptability of a Lifelogging Wearable Camera in Older Adults with Mild Cognitive Impairment: A Mixed-Method Study, BMC Geriatrics, vol. 19, Article 110 (2019).
[6] K Takahashi, D Yokoyama: Remembrance Support, Journal of the Robotics Society of Japan, vol. 28, no. 9, pp. 1082–1083 (2010) (In Japanese).
[7] T Yagi, T Nishiyasu, K Kawasaki, M Matsuki, H Sato: Go-Finder: A Registration-Free Wearable System for Assisting Users in Finding Lost Hand-Held Objects, IPSJ Interaction 2021 Proceedings, pp. 739–744 (Mar. 2021).
[8] Nextcloud: https://nextcloud.com/ (Visited on 2024-12-19).
[9] PhotoPrism: https://www.photoprism.app/ (Visited on 2024-12-19).
[10] Raspberry Pi Zero 2W: https://www.raspberrypi.com/products/raspberry-pi-zero-2-w/ (Visited on 2024-12-19).
[11] Raspberry Pi Camera Module 3: https://www.raspberrypi.com/products/camera-module-3/ (Visited on 2024-12-19).
[12] MySQL: https://www.mysql.com/jp/ (Visited on 2025-08-05).