Master's Thesis

LocatAR: An AR Object Search Assistance System for a Shared Space

Hiroto Oshimi
Program of Information Science and Engineering
Graduate School of Science and Technology
Nara Institute of Science and Technology

Supervisor: Prof. Kiyoshi Kiyokawa
Cybernetics and Reality Engineering Lab (Division of Information Science)

Submitted on January 31, 2023

A Master's Thesis
submitted to Graduate School of Science and Technology,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
Master of Engineering

Hiroto Oshimi

Thesis Committee:

Supervisor	Kiyoshi Kiyokawa (Professor, Division of Information Science)
Co-supervisor	Keiichi Yasumoto (Professor, Division of Information Science)
Co-supervisor	Hideki Uchiyama (Associate Professor, Division of Information Science)
Co-supervisor	Naoya Isoyama (Assistant Professor, Division of Information Science)
Co-supervisor	Monica Perusquía-Hernández (Assistant Professor, Division of Information Science)

LocatAR: An AR Object Search Assistance System for a Shared Space*

Hiroto Oshimi

Abstract

Item-finding tasks due to memory lapse are costly activities commonly experienced by many people. However, item pre-registration, privacy protection for multi-users, and item search methodology have various issues left unresolved. Therefore, we propose a multi-functional, pre-registration-free, and 3D location-based item management system. The system has two main functions: registration and search. The automatic registration is performed by image-based item movement recognition from the user's grasping and placing motions. The registered item movement data comprises the item category, and the start and end locations. We ensure privacy protection by storing item movement data without images. Also, we provide a user interaction to refuse to share the items with other users. The search is based on the item list or item location. The location-based search is performed by specifying where the user last saw the item. To optimize and test the performance of the system, we performed two studies: parameter optimization and a search experiment. The parameter optimization performed in the auto-registration experiment led to the discovery of optimal values that are difficult to reach empirically. The search experiment showed that the proposed system's search and guidance functions are effective as an assistance system for finding items, both in terms of search time and user experience. Overall, our system demonstrated the potential to be a useful assistance system for managing items in a shared space. We further discuss the possibility of further exploiting the limited registered information by treating item location as an identifier of the moved item.

*Master's Thesis, Graduate School of Science and Technology
Nara Institute of Science and Technology, January 31, 2023.

Keywords:

Object Finding, Augmented Reality, computer-aided tools

Contents

List of Figures

1 Introduction

In spaces such as offices, laboratories, factories, etc., where multiple people cooperate or perform their individual tasks (hereinafter referred to as "cooperative environments"), shared equipment and fixtures are often placed (henceforth, these items are referred to as "items"). Unless these items are under strict control, they can be left by users in places other than where they are supposed to be. This causes a search task when other users need the item. Item search is an unexpected event that is time-consuming. Search tasks occur because the searcher does not know the location of the item. If the searchers themselves are the previous user, they can limit the scope of the search by looking back at their own actions. For others, the cost of the item search task may be higher. Therefore, search for shared items in a cooperative space is more difficult than search for personal items in personal areas. Cooperative spaces have a large number of shared items, placed in a wide search area, and many users move the items.

Item search assistance systems are related to the technical fields of recognition and navigation: recognition of item location, user behavior, and user-item interaction, as well as the guidance of the user by the system, each of which has been studied for a long time. Various methods have been proposed for guiding users. However, like yagi's go-finder [1], which provided users with images as a timeline of movement, many of the query methods are based on a method where the user naïvely searches for the desired move in a full search from all moves. It requires users spend much time to query for movement. Therefore, an efficient method is desired. Compared with the existing methods, we propose to use the user's knowledge of the item and the locations where the item should be or was previously located as a query to narrow down the candidates of movement. In the item search task, Augmented Reality (AR)-based methods have clear advantages over other methods. It is easy for the users to know the relative position between the item to be searched for and themselves. Also, the system can continuously guide the user until the user finds the item. In our proposed method, item location is automatically registered. Therefore, our system will exploit these advantages as an AR system using the AR Head-Mounted Display (HMD) HoloLens 2.

Privacy in multi-user use is another important issue for conventional methods of finding items support. Many IoT systems perform automatic movement collection using such technologies as cameras and RFID [1, 6]. When the system automatically registers items and movements, such information may inadvertently violate the privacy of other users. Therefore, we focus on an item search support system that solves these problems in movement registration and user guidance, and describe its implementation and evaluation.

We propose to design, implement and evaluate a system to assist in the search for shared items in a cooperative space. The system enables effective guidance using location by automatically acquiring the timing of item movement and its location at that time. Our proposed search system will assist users in finding things by guiding them with the minimum necessary information related to their movement for the protection of the privacy of all users. To accomplish this in a way that eliminates the cost of prior item registration, we design a user interaction and user interface (UI) that allow the user to refuse registration of movement. Two novel technical ideas are the recognition of the beginning and end of the movement of the grasped item in real-time from first person view image, based on item recognition and optical flow prediction, and a search system that leverages the user's knowledge of the item's previous location. The two research questions of this study are: (1) is it possible to recognize the start and end of the movement of a grasped item from first-person images in real-time using item recognition and optical flow prediction?; and (2) can location-based search functions and visualization of movement without using images be an efficient search aid for the user?

Our contributions include:

- Real-time item grasp and movement recognition technology without pre-registration of items using optical flow.
- Location-based search UI to help users find items.

- Location-based movement visualization to guide users.

2 Related works

2.1 Item Locating

In 2013, Funk et al. introduced a system for managing a place of registered real-world items, which enabled users to find the places where the items are by using a stationary monitor [7]. By using a motion tracking system, it can track a web camera on the user's head. There are also methods using such as RFID tags [8–13] or Bluetooth [14–16] that provided a highly accurate item location. However, in environments where a large number of items are managed or frequently replaced, the time and monetary cost of the pre-registration work is a burden on the user. Therefore, it is desirable to eliminate the need for pre-registration of such items. Farasin et al. made an item-tracking method using Microsoft HoloLens [17]. Their method used the HoloLens camera and position tracking to enable recognition of the category and 3D location of the item without prior registration. We have adopted this object position recognition method as part of our object grasp and movement recognition system.

2.2 Item Recognition

Regarding the recognition of items, many methods were studied that track pre-registered items. For recognition of user interaction with items, Maekawa et al. proposed a method to recognize user actions using multiple types of wrist-mounted sensors [18]. Wu et al. recognized hand posture from images taken by a wrist-mounted camera to recognize grasping and other actions [19].Ueoka et al. have proposed a system called "I'm Here!" that identified items in an image by extracting a histogram of pixels from the area of the item in the camera image and comparing it with pre-registered data [20]. Further, Yagi et al. designed "GO-

Finder" in which a camera installed on the user's chest recognized the interaction state between the user's hand and an item and recorded the camera image at that moment [1]. Their method eliminated the burden on the user to pre-register items to use the system, which had been a problem with earlier item search assistance systems.

2.3 Search Assistance Systems

For user guidance, many methods presented images and the item's current location. Many wearable systems that present camera images to the user are intended to assist the user's memory [6]. "I'm Here!" by Ueoka et al. assisted users in their search by presenting camera videos of a certain time period centered on the point when the item was last viewed [20]. Another system, "GO-Finder" enabled users to look for an item by presenting a list of images displaying the moment of grasping within a smartphone application [1] as shown in Fig.2.1.

The use of item search assistance systems in spaces where multiple people are present has been envisioned in recent years. Yan et al. [2] recorded people who were in the same location as the item from the camera image, or who were frequently with the item, and created an option for users to ask these people about the location of the item when search for it as shown in Fig.2.2. In the figure, 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. This allowed assistance in the search even if the item itself was not captured by the camera. However, privacy issues were discussed because of the registration of images containing other people's faces. Hoyle et al. [21] noted that many users have privacy concerns about sharing and storing images using a lifelogging system with a wearable camera.

Many studies on item finding systems only mentioned privacy issues in the discussion but did not solve them. "LiSee" by Chen et al. [4] is an example of an assistance system for blind and low-vision (BLV) people as shown in Fig.2.4 that solves the privacy issue as a side effect. While most conventional assistance systems for BLV users showed camera images of the user's daily life to assistants, which raised privacy concerns, this system used voice to assist BLV users in finding things on their own. However, guiding the user by voice, as LiSee has done, is considered unsuitable for the purpose of selecting the user's desired movement from a large amount of movement information. This is because it is difficult for the user to process multiple voices at once, which is a considerable disadvantage compared to a visual display such as a list of images.

Nakada et al. showed users the location of the item they were looking for by highlighting its location with a spotlight installed on the ceiling [3] shown as in Fig.2.3. The advantage of the item location highlighting method in real space is that the use of real space, which is easily recognized by the user, allows for guidance with fewer elements of misunderstanding or confusion. The disadvantage, however, is that the limitations of real space can cause problems, such as when the movement of an item crosses multiple spaces and the use of spotlights in a cooperative environment interferes with the work of others. As a guidance method that utilizes only this advantage, we believe that it is desirable to use a method like AR, which is essentially the same as presentation in real space from the user's perspective and is not subject to the constraints of real space or other people.

Funk et al. investigated the optimal display of an item's location using AR in a search task [5]. They compared two methods: (1) a display of the user and item locations in a two-dimensional map; (2) an arrow display indicating distance and direction at remote locations and a camera image display of the moment when the user last saw the item when the user approaches the item to a point five meters away from the item as shown in Fig. 2.5. They concluded that the method that displayed arrows and camera images was superior. In their method, the presentation of images compensated for positional inaccuracies. Therefore, we think that the images should not be necessary when a more accurate location can be available. Furthermore, image presentation has the problem that viewing angles and obstructions can make finding the item difficult. Our method uses accurate item location acquisition and user tracking to present accurate locations to the user.

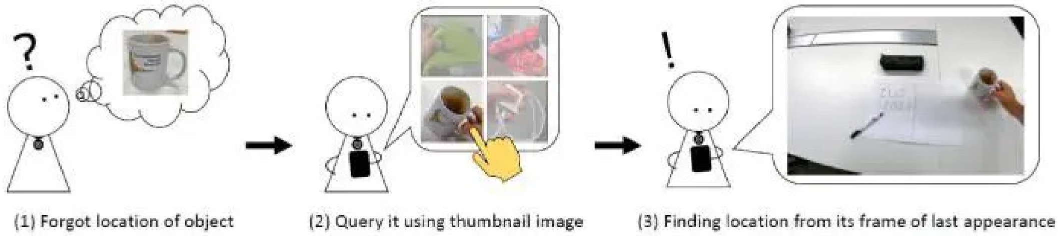

?
!
(1) Forgot location of object
(2) Query it using thumbnail image
(3) Finding location from its frame of last appearance

Figure 2.1: An object finding assistance system which presents a list of images of the moment an item is grasped to the user [1].

2.4 Position of our research

We believe that a wearable AR system capable of presenting information to a user who is walking around is highly compatible with the item search task. Since the 1990s, research on the use of AR for information presentation and guidance methods has continued to be proposed until now. In 1997, Starner et al. formed a community of wearable computer users and introduced their AR wearable computing systems [22]. Among them, Rhodes made a wearable remembrance agent which reminded the user of schedules and other resources which are related to the current context [23]. Suomela et al. introduced a taxonomy for methods of visualizing location-based information, which used an environment model and user's view perspective to analyze visualizing UI [24]. According to the analysis, one of the merits of the 3D environment model is that it does not demand users excessive matching between places in real and virtual environments.

Darken et al. analyzed real-world travel, in their paper called wayfinding [25]. They applied environmental design principles to virtual environments. In their analysis, there were three types of search tasks; naïve search, primed search, and exploration. These tasks can be compounded into sequences. In this manner, users tend to conduct a naïve search first, followed by a primed search. In contrast to virtual reality (VR), AR enables users to work in a known environment around them. Therefore, we expect the item search task to be a pure primed search. In 1997, Azuma described several AR applications which will be possible when high-accuracy HMD tracking is accomplished [26]. One of them is a visualization of object locations. The system we have created belongs to "annotation and

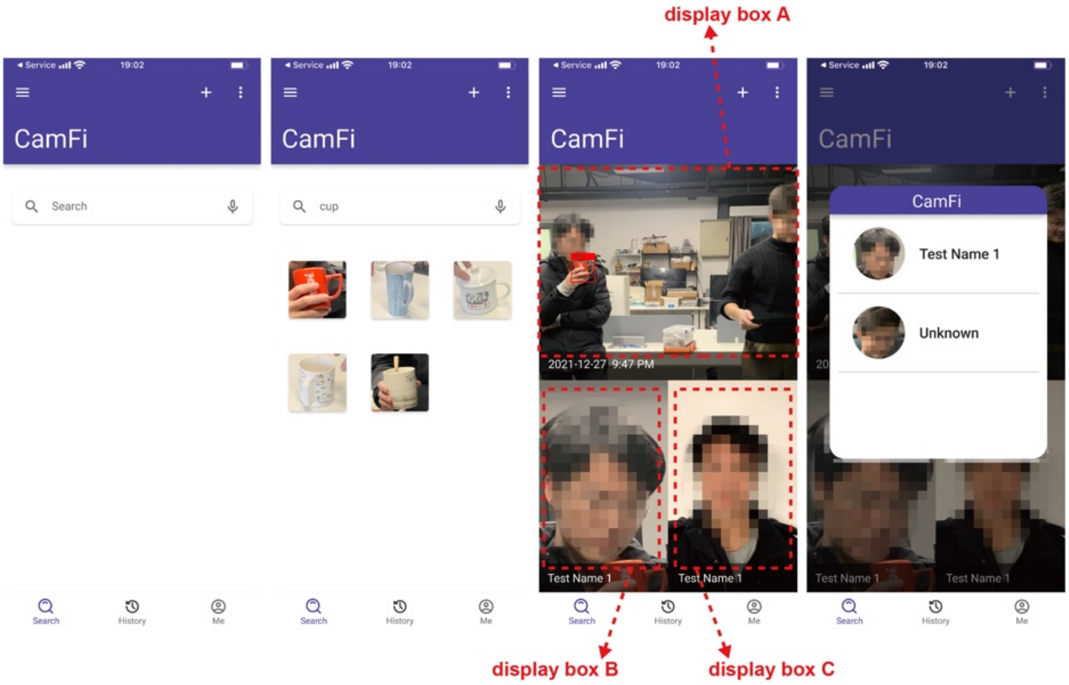


Figure 2.2: The assistance system for finding items using a person in the frame at the same time as the item [2].

visualization" in Azuma's six classifications of AR applications.

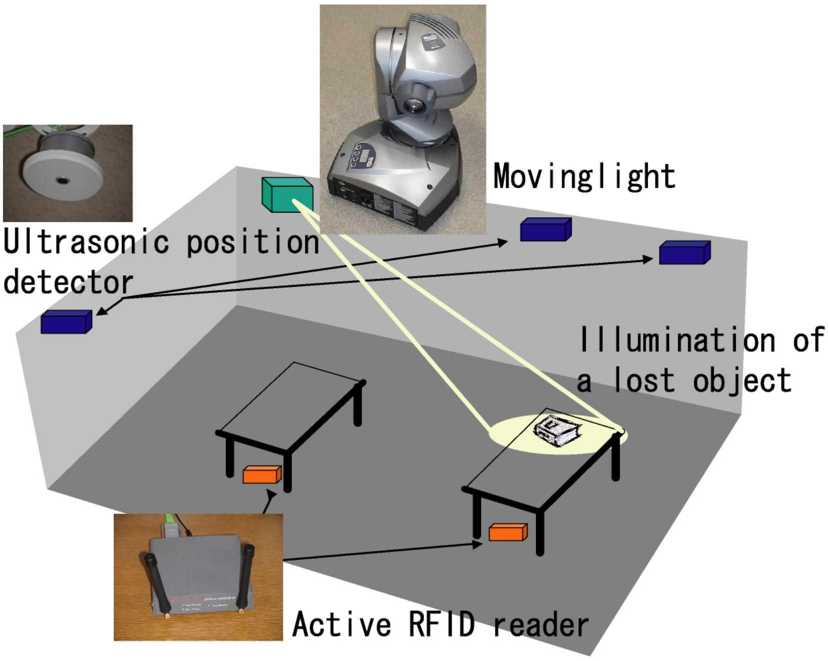


Figure 2.3: The item location indication system using RFID and a spotlight [3].

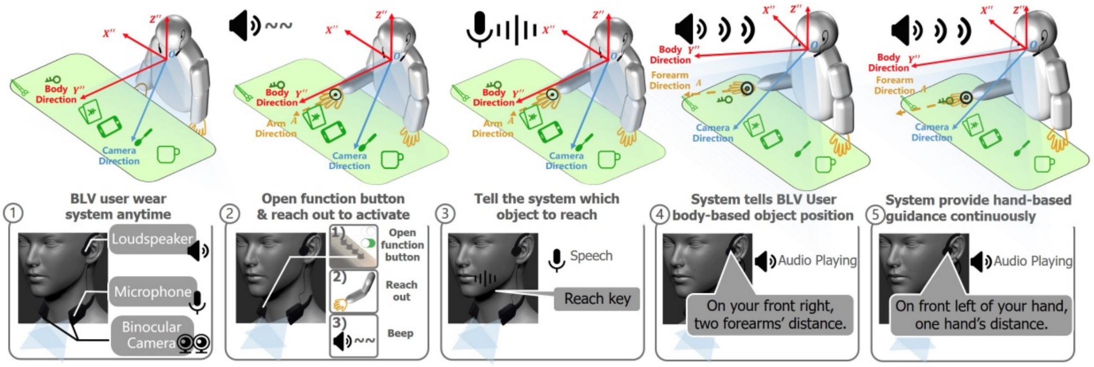

X''
Z''
Y''
Body Direction
Arm Direction
Forearm Direction
Camera Direction
1 BLV user wear system anytime
Loudspeaker
Microphone
Binocular Camera
2 Open function button & reach out to activate
1) Open function button
2) Reach out
3) Beep
3 Tell the system which object to reach
Speech
Reach key
4 System tells BLV User body-based object position
Audio Playing
On your front right, two forearms' distance.
5 System provide hand-based guidance continuously
Audio Playing
On front left of your hand, one hand's distance.

Figure 2.4: The item location indication system using a headphone [4].

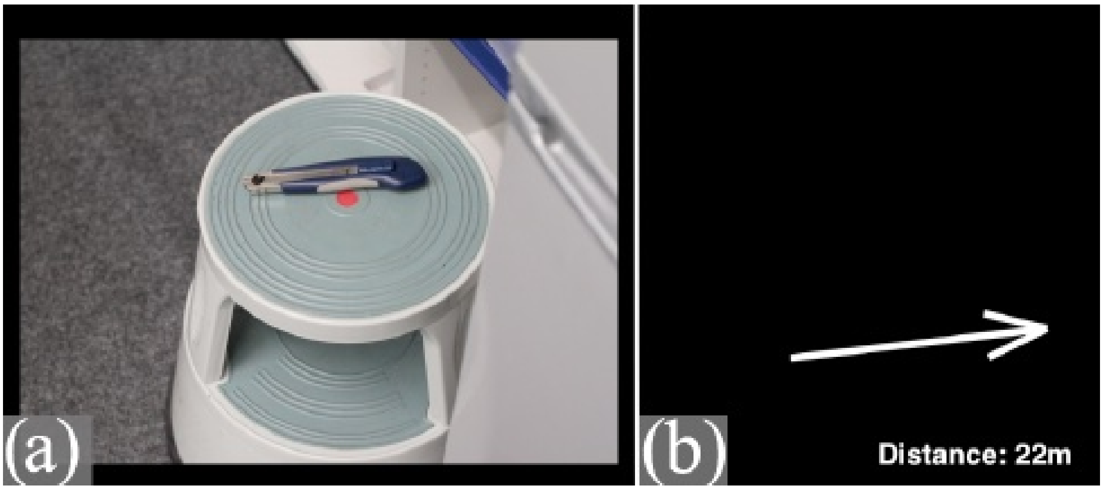


Figure 2.5: The last-seen image representation on the glasses. (a) Lastseen image of the sought object. (b) Arrow pointing towards the target when the user is more than 5 meters away [5]

3 Proposed System

The proposed system assists a user's lost item finding by visualizing the item's movement as shown in Fig. 3.1. Our system has two main functions: registration and search. The registration function monitors user activity and registers moved items as shown in Fig. 3.2 while the search function visualizes 3D item movements on the HMD. The registration function also provides the user interaction to refuse to share moved items with others for privacy protection.

To achieve these functions, the proposed system uses an AR HMD and a processing server. The AR HMD and the processing server communicate as a client-server system. The AR HMD has a camera to send camera images to the server and visualize the item data provided by the server. On the server, item movement recognition and its data management are performed. We adopted Microsoft HoloLens 2 as our system's AR HMD.

3.1 Registration

The system automatically registers items by detecting the user's item grasping and item placement from the camera images. The registered item data comprises the item category, its 3D positions and clock times before the grasping and after the placement.

As shown in Fig. 3.2, the HMD is constantly acquiring frames while the system is in use, and the acquired frames are sent to the recognition server at the timing when the processing of past frames in the HMD is completed. The recognition server estimates the category and item moving state which indicate a moving state of the items in the received frames, and sends the results to the HMD. The HMD receives the recognition results and determines the item state, which indicates which stage of the grasping and moving event has occurred for each

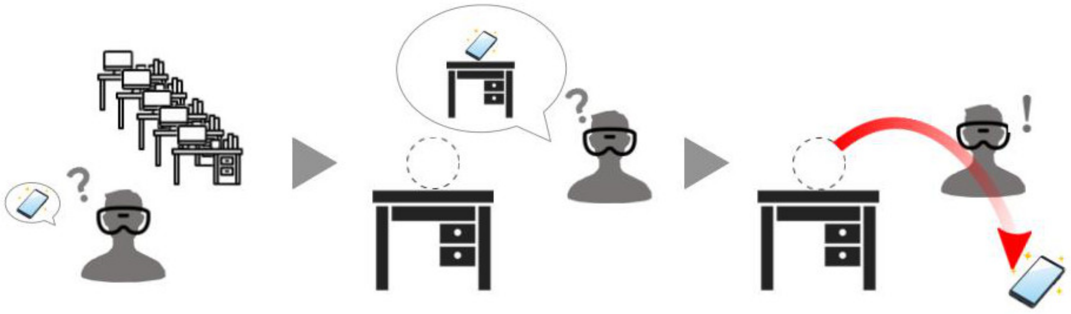


Figure 3.1: sytem overview.

item in the recognition results. The user state, which indicates that the user is grasping or placing the item, changes according to the determined item state.

The system notifies the detection event to the user at the beginning and the end of the item movement. The user can correct it on the 3D user interface if the detection event is wrong, as illustrated in Fig. 3.3. The same procedure applies when a user does not want to share an item movement. In this case, the user can use this interaction to deny item movement sharing with others. The interaction detail is as follows. When item grasping is detected, the system immediately notifies the user by showing a cube around the object with text describing the recognized object category (Fig. 3.3 (a)). The user can revoke the detection result by pressing the revoke button on the UI if necessary. After grasping, the hand or the item does not need to appear in the camera frame and the user can hold and move it with natural posture to keep the grasping state. When the item placement is detected, the system also notifies the user (Fig. 3.3 (b)). At this moment, the rejection button is displayed in additoin to the revoke button. The user can choose whether the item will be registered in the system or not. The user can revoke the placement by the revoke button. In this case, the system returns to the state where the user is still grasping the item (Fig. 3.3 (c)). On the other hand, the entire item movement will be removed if the user presses the rejection button, and the system returns to the state where the user is grasping nothing (Fig. 3.3 (d)). The system registers the item movement only when the user does nothing for a certain period.

When registering the items, their 3D coordinates are also stored. On the

HMD, both visual SLAM and dense scene reconstruction are performed owing to the HMD functionality. After detecting the items on the server, their 2D screen coordinates are sent to the HMD. Then, their 3D coordinate is computed by raycasting, exploiting the densely-reconstructed scene. Finally, the 3D item coordinates are registered on the server.

There are two main reasons for using the interaction between the user and the item as the trigger for registration. First, the system records the start and end of item movement because the same user moves the item between specific locations. This is generally essential information for search objects. Second, the system asks the user whether the item should be registered at the end of the movement interactively. This is because it would be difficult for the user to judge it when time has passed since the end of the movement. Therefore, the system is designed to interact with the user right after the placement.

3.2 Search

The system visualizes 3D item movement with a curved line in AR, as illustrated in Fig. 3.4. The curve ends correspond to the grasping location (in red) and the placement location (in blue) of the item. This 3D visualization can effectively guide the user from the grasping location to the placement location, even in a large space. To make the newer curves more noticeable, the thickness of the curves is adjusted. The line thickness is inversely proportional to the time that has passed since the item moved. Also, to avoid overlapping curves in the same direction, the height of the curve is proportional to the distance the item travels.

This system offers two types of item search methods for the user. The first method is to select an item from the registration list, as illustrated in Fig. 3.5. The user checks all the items in the list to find the item of concern. The list is generated in order of newest to oldest registered items. Each item data contains the category, the grasping and placement times, and the 3D coordinates. The item movement is visualized in AR when the user selects the item. The second method is location-based search, as illustrated in Fig. 3.6. Fig. 3.7 is the explanation of the whole activity of this search function. After the user selects location-based search, HMD accepts the user's specification of the location, and HMD retrieves all the movements recorded by the system at that time from the server. The HMD then uses the specified location to narrow down the candidates for the user's desired item movement. All the candidates obtained as a result of the narrowing down are visualized, and the user can confirm the details of the movement by selecting one of them and can highlight the location where the movement ended. As the user interface, the specification of location is accomplished by raycasting from the finger to the reconstructed surface. We especially use Air Tap provided by HoloLens 2 for this process. After computing the location, the items near the intersection are retrieved. Finally, the user selects one of them to visualize its detail. This method is useful for the user when finding an item at the location where it was last found. We reduce the number of candidates the user checks by utilizing their knowledge of where the lost items were located.

For location-based search, an interactive sphere is placed on the curve. This sphere provides the function of displaying the movement detail and highlighting the endpoint. When the user touches the sphere, the system shows the movement detail in the indicator on the HMD screen. The system also highlights the curve endpoint, allowing the user to find the placement location easily.

These search methods limit the item candidates to be visualized because of two reasons. The first one is that the greater the number of item movements that are visualized, the greater the effort to look identify the desired item movement from the available choices. The second one is that the intersection of the visualized curves can be an obstacle to the user's line trace.

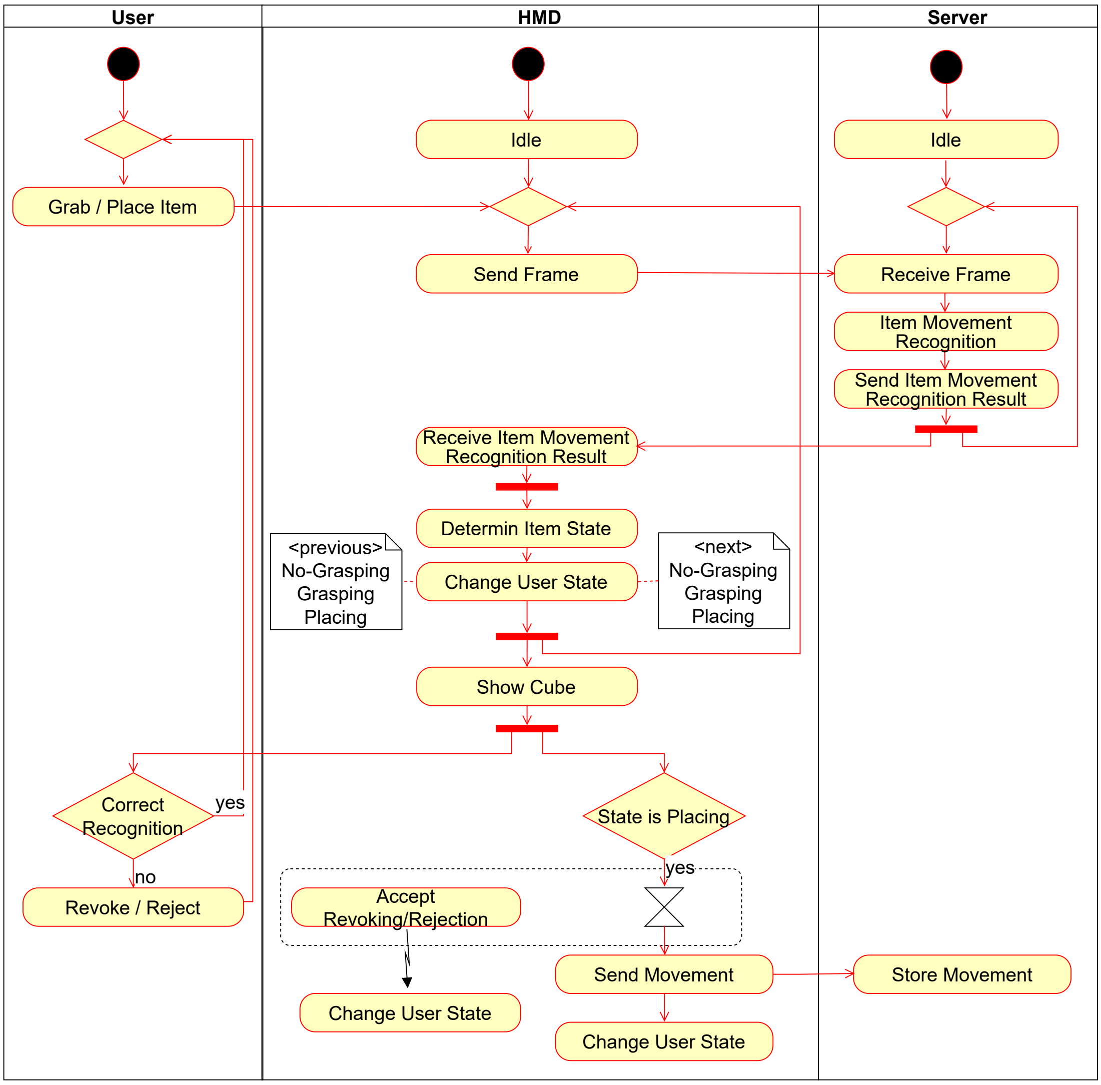
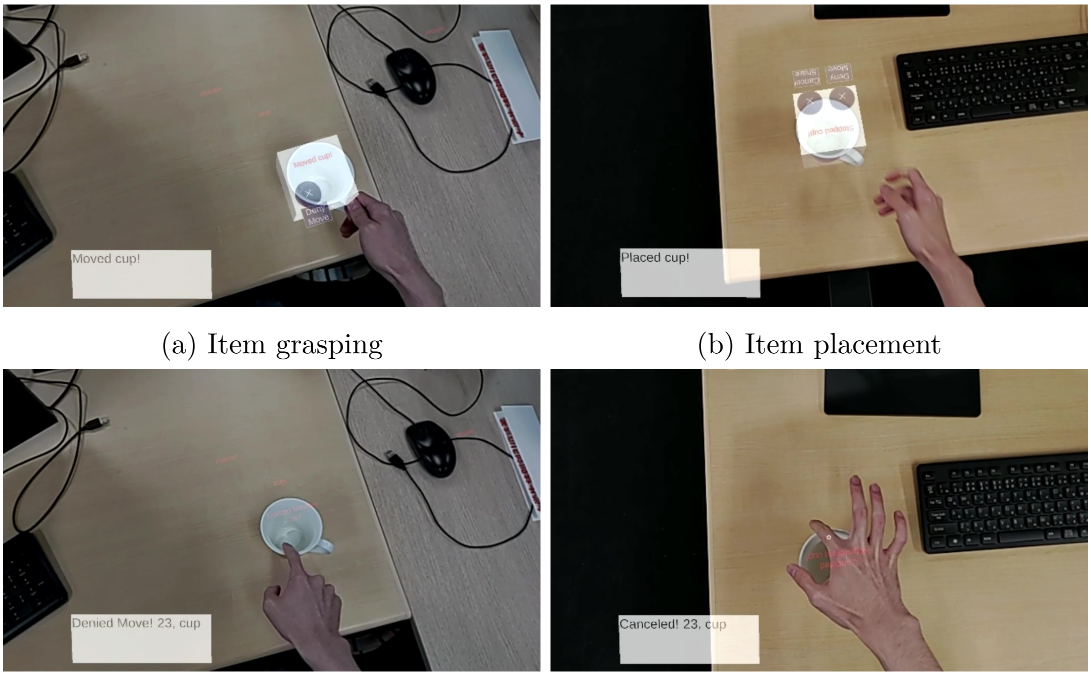


Figure 3.2: Registration flowchart.



(a) Item grasping

(b) Item placement

(c) Item revoke

(d) Registration rejection

Figure 3.3: Item movement recognition and user interaction.

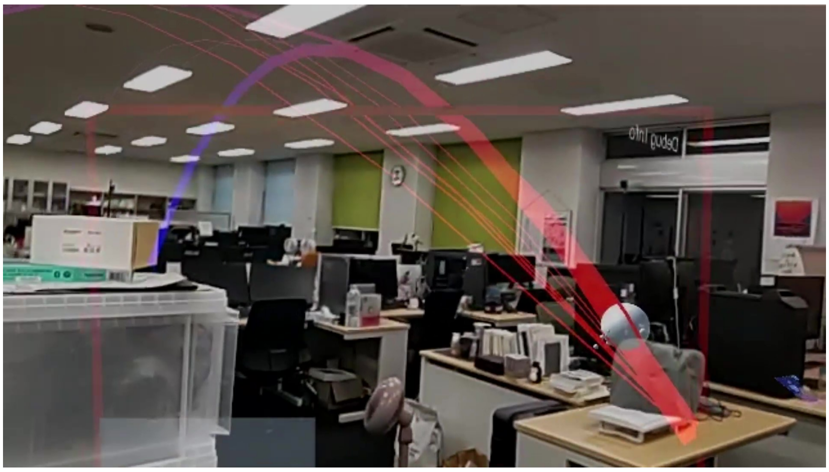
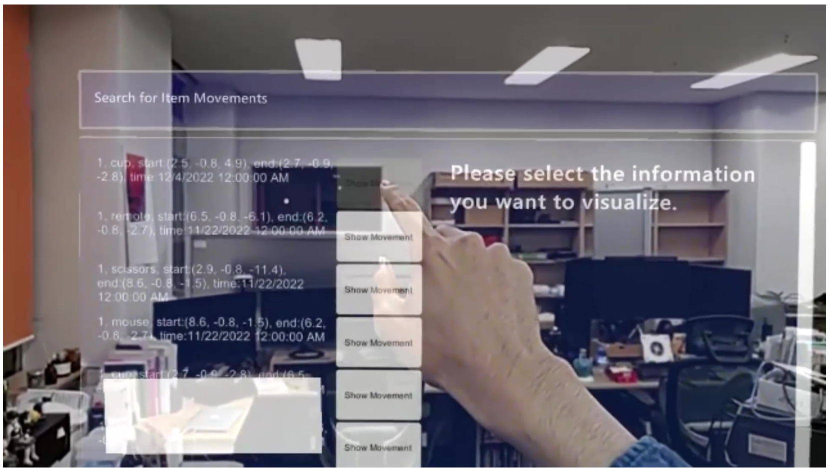


Figure 3.4: Item movement visualization in AR.



Search for Item Movements

1, cup, start:(2.5, -0.8, 4.9), end:(2.7, -0.9, -2.8), time:12/4/2022 12:00:00 AM

1, remote, start:(6.5, -0.8, -6.1), end:(6.2, -0.8, -2.7), time:11/22/2022 12:00:00 AM

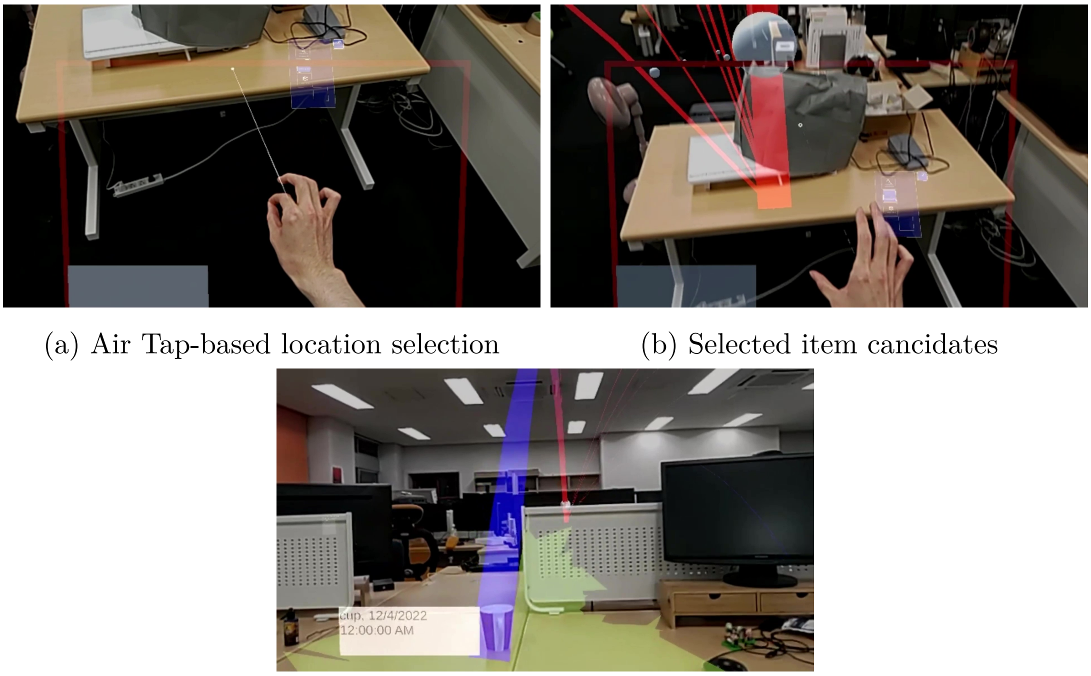
1, scissors, start:(2.9, -0.8, -11.4), end:(8.6, -0.8, -1.5), time:11/22/2022 12:00:00 AM

1, mouse, start:(8.6, -0.8, -1.5), end:(6.2, -0.8, -2.7), time:11/22/2022 12:00:00 AM

1, cup, start:(2.7, -0.9, -2.8), end:(6.5,

Please select the information you want to visualize.

Figure 3.5: Item list-based search.



(a) Air Tap-based location selection

(b) Selected item cancidates

(c) Highlighted placement location

Figure 3.6: Location-based search.

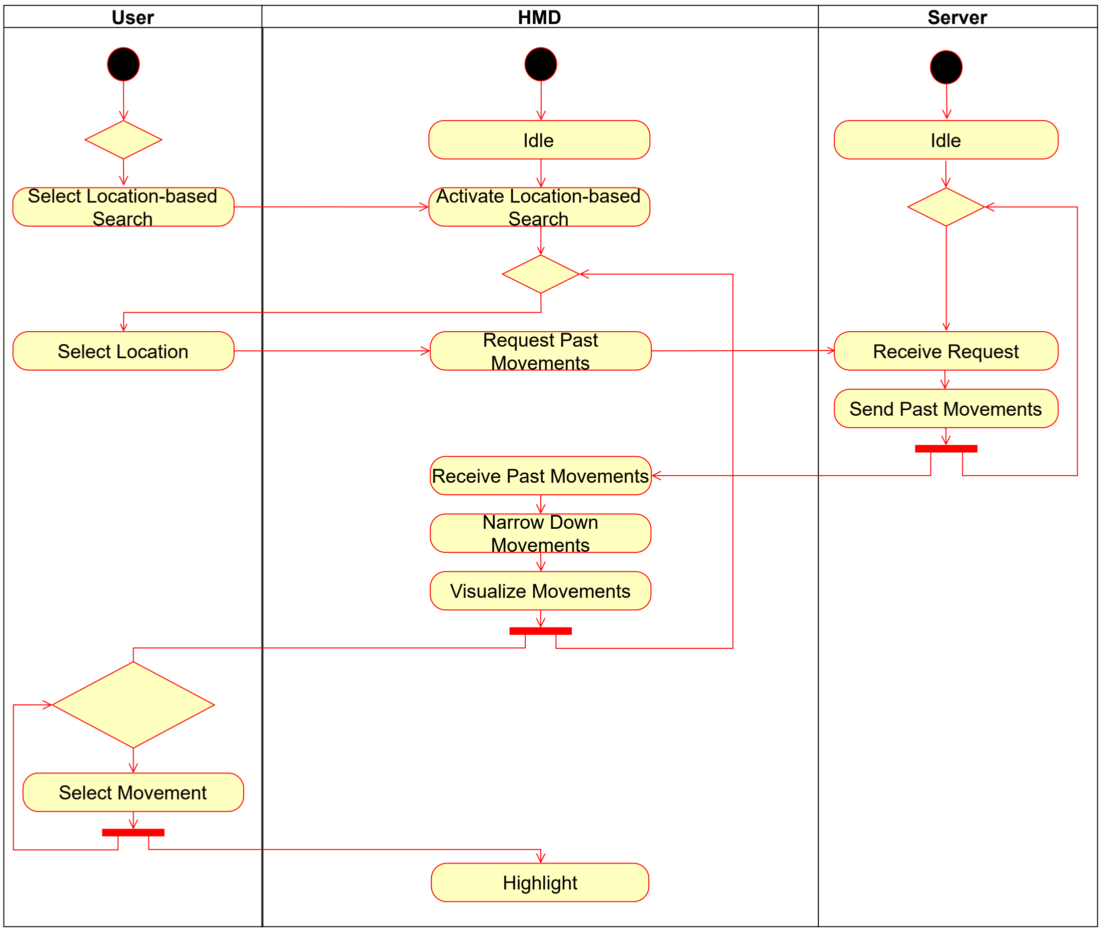


Figure 3.7: Location-based search flowchart.

4 Algorithm

4.1 Registration

Our system realizes the registration process through the management of item movement state, item state, and user state. The item movement state indicates whether the item is moving or not, and the item state has the state that changes during the movement phase of the corresponding item. The user state indicates the stage of the movement that the user is performing at that moment. In this section, we explain how to manage the user state in 4.1.2 and determine the item state and item movement state in 4.1.1. Finally, we describe the storing of movements in 4.1.3. These subsections refer the steps in 4.1.

4.1.1 Item State Determination

In the item registration, the processing server performs item recognition and item-movement detection while the HMD computes 3D item coordinates and verifies them. The item is registered only when the movement event detected by the server passes the verification using 3D hand and item positions.

Our method is based on an existing method [27], as illustrated in Fig. 4.2. This method uses the optical flow and the distance between the item and the hand for determining the item status such as a new item, grasped and placed. These status means as follows.

- grasped: The item was not grasped.
- placed: The item was not placed.
- new item: The item was a new item.

The recognition server estimates an item's movement status such as static and moving. These status means as follows.

- static: The item has not been moved.
- moving: The item has been grasped and moved.

YOLACT++ [28] and RAFT [29] are first applied to each image to detect the items and hands and compute their flow. The average of the optical flow in each area is computed, from the item's optical flow ($\boldsymbol{F}_i$), hand's optical flow ($\boldsymbol{F}_h$), and background optical flow ($\boldsymbol{F}_{bg}$). YOLACT++ simultaneously provides the category information of the item. We check for any discrepancies in newly grasped or placed items by comparing the category (C_g) of the most recently grasped item with the category (C_t) of the target item of the current process. Also, the 3D item coordinate ($\boldsymbol{P}_i$) and 3D hand coordinate ($\boldsymbol{P}_h$) are computed. The aforementioned data is used together with the following criteria to determine the item status as shown in 4.3. The system recognizes that the user has grasped an item if all of the following equations (4.1), (4.2) and (4.3) are satisfied.

$$|\boldsymbol{F}_i - \boldsymbol{F}_{bg}| - |\boldsymbol{F}i - \boldsymbol{F}_h| > T * ElapsedTime \tag{4.1}$$

$$C_g = NONE \tag{4.2}$$

$$T_l < |\boldsymbol{P}_i - \boldsymbol{P}_h| < T_h \tag{4.3}$$

where T is a parameter to be optimized in the evaluation, $ElapsedTime$ is the time interval between the acquired time of the image being processed and the one previously processed, and T_l and T_h are low and high thresholds for 3D distance between the hand and each item. $NONE$ means no item has been grasped. Alternatively, the system recognizes that the item is deemed static if the equations (4.3) above and (4.4) below are satisfied.

$$C_g = C_t \tag{4.4}$$

In the move condition, T_l and T_h are set to 0 [m] and 0.4 [m], respectively. In the place condition, T_l and T_h are set to 0.1 [m] and 0.4 [m], respectively. Note that in the current implementation, the registration database resides on the HoloLens 2 to ensure the maximum performance for the image processing on the recognition server, thus it does not support multiple users in reality.

4.1.2 User State Management

The system maintains a user state that indicates whether the user is grasping an item, and changes the processing of information about the item obtained from the recognition server according to the user state of no-grasping, grasping, and placing. This user state indicates as follows.

- no-grasping: The user is not grasping items.
- grasping: The user has an item in his/her hand.
- placing: The user has placed an item right before.

A user state changes as shown in Fig. 4.4. If a grasp occurs when the state is no-grasping, the state changes to grasping; if a place occurs when the state is grasping, the state changes to placing. When a certain period of time elapses when the state is placing, the state changes to no-grasping, and the move will be registered at that time. The UI for revoking and rejection is achieved by changing this state. Revoking takes place between grasping and placing states, and when a revoking occurs, the state goes back to the previous one. When a rejection occurs, the placing moves to no-grasping without registering any movement information. In other words, this is equivalent to two revoking occurring at once.

4.1.3 Movement Storing

Information on the following aspects of movement is stored as shown in Fig. 4.5.

- StartPosition: 3D coordinates of the place where a user grasps an item.
- EndPosition: 3D coordinates of the place where a user grasps an item.
- Category: The category to which the item belongs.
- ItemID: A number that identifies the item (not yet implemented to be unique).
- Date: The time when the item is repositioned and the start and end of the move are both confirmed.

The location of an item obtained from 2D to 3D location conversion is a relative position based on the point from which the system was launched. When registering the movement start/end locations, it is necessary to register the relative position from the unified base location. In the implementation, the user makes the system recognize a fixed QR code at launch. When registering the move, this QR code is used as the unified reference point, and the location information is converted from the coordinate space based on the system launch location to the coordinate space based on the QR code. The category of the item is the recognition result of YOLACT++ trained on the COCO dataset. The ID is assigned to the item with a number that increases by one in the order in which the item is newly recognized by the system. The date is registered as the time when the waiting time for revoking or rejection by the user ends after the item is placed. In the conceptual diagram of the system, the movements sent by the HMD are stored in a database on the server side, but in the actual implementation, they are managed in text files on the HMD side. This is a result of consideration of ease of operation checks in implementation.

4.2 Search

4.2.1 Visualization

In the visualization, the system generates a curve in real-time. The curves are Bézier curves with endpoints at the start and end points, and their curvature increases with the distance of the corresponding movement. The purpose of this is to reduce the overlap of multiple movements when they are visualized, and to reduce the overlap with real obstacles when visualizing long-distance movements. The Bézier curve is drawn with three control points, two indicating the start and end points ($\boldsymbol{P}_S$, $\boldsymbol{P}_E$) and one adjusting the height between them ($\boldsymbol{P}_C$). $\boldsymbol{X}_C$, $\boldsymbol{Y}_C$ and $\boldsymbol{Z}_C$ are calculated by Eq.4.5 through 4.7. By using these three points, the system visualizes a movement as shown in 4.6.

$$\begin{aligned}
\boldsymbol{X}_C &= (\boldsymbol{X}_S + \boldsymbol{X}_E)/2 + \boldsymbol{C}_H * log(1 + |\boldsymbol{P}_S - \boldsymbol{P}_E|) && (4.5)\\
\boldsymbol{Y}_C &= (\boldsymbol{Y}_S + \boldsymbol{Y}_E)/2 && (4.6)\\
\boldsymbol{Z}_C &= (\boldsymbol{Z}_S + \boldsymbol{Z}_E)/2 && (4.7)
\end{aligned}$$

where $\boldsymbol{C}_H$ is a constant for height adjustment.

The newer the curve, the thicker it is displayed. This is due to our idea that the moves that users look for more frequently in their search are newer moves and should be emphasized to them. The thickness of the curve is inversely proportional to the number of days ($\boldsymbol{D}$) elapsed, as in Eq. 4.8.

$$\boldsymbol{W}_L = \boldsymbol{C}_W/\boldsymbol{D} \tag{4.8}$$

where $\boldsymbol{C}_W$ is a constant for the width adjustment.

4.2.2 Narrowing-Down the Movements by Location

As an assistance system for finding the move the user is looking for, the system performs a move filtering function. We expect our system to reduce the user's cost of search for a move by providing a location-based search function. The system uses the coordinates of a user-specified location and past movement information to narrow down moves that started near the user's specified location.

In location-based search, the system first receives the user's specification of the search location and obtains the coordinates ($\boldsymbol{P}_U$) of the location based on the system's starting point. Next, the system retrieves all previous movements from the database and obtains the StartPosition ($\boldsymbol{P}_S$) and EndPosition of each movement. Since these Positions are stored as coordinates of the QR code-based coordinate system, the system converts them to the same coordinates of the system start point-based coordinate system as $\boldsymbol{P}_U$. As shown in Eq4.9, the system presents to the user only those moves for which the distance between $\boldsymbol{P}_U$ and the transformed $\boldsymbol{P}_S$ ($ToLaunchCoordinate(\boldsymbol{P}_S)$) is less than a threshold value as candidates for the move the user is seeking.

$$|\boldsymbol{P}_U - ToLaunchCoordinate(\boldsymbol{P}_S)| < T_h \tag{4.9}$$

4.2.3 Highlighting

Highlighting is a function that provides a more detailed visualization of a movement, and when a curve is selected, it displays detailed information in an indicator and a spotlight around the end point of the movement. The indicator displays the category and date of the moved item, information that cannot be accurately retrieved from the visualization of the movement alone. The highlighting representation uses the location of the endpoint of the movement and a 3D mesh of the place. When a user select visualized information, the system changes the color of polygons within a certain distance from the endpoint to yellow-green to achieve the highlighting expression.

In our actual system, we use spatial mesh information obtained by the HoloLens 2 functionality as the 3D mesh of the place. Since the mesh is periodically updated, the system can adapt to changes in the surrounding environment. In the implementation, a variable in the shader attached to the spatial mesh is updated on the event of movement information selection. This variable holds the position of the end point, and the color is recalculated on the next frame after the variable is updated.

4.2.4 Selection

In order for the user to select the desired movement from among multiple visualized movements, a UI for selection is needed. To prevent user confusion, instead of using a method where the user takes his/her eyes off the displayed curve and looks at the UI elsewhere, a method where the user interacts with the curve presented to him/her is used. One intuitive way to achieve this is to highlight the curve by touching the curve itself. However, this visualization method, which generates curves in real-time, requires a high computational cost for the generation of the curve colliders. We employ a method in which a sphere is placed at the root of the starting side of the curve, and the highlight is executed when the user touches it. Therefore, we give the collider to the sphere instead of the curve.

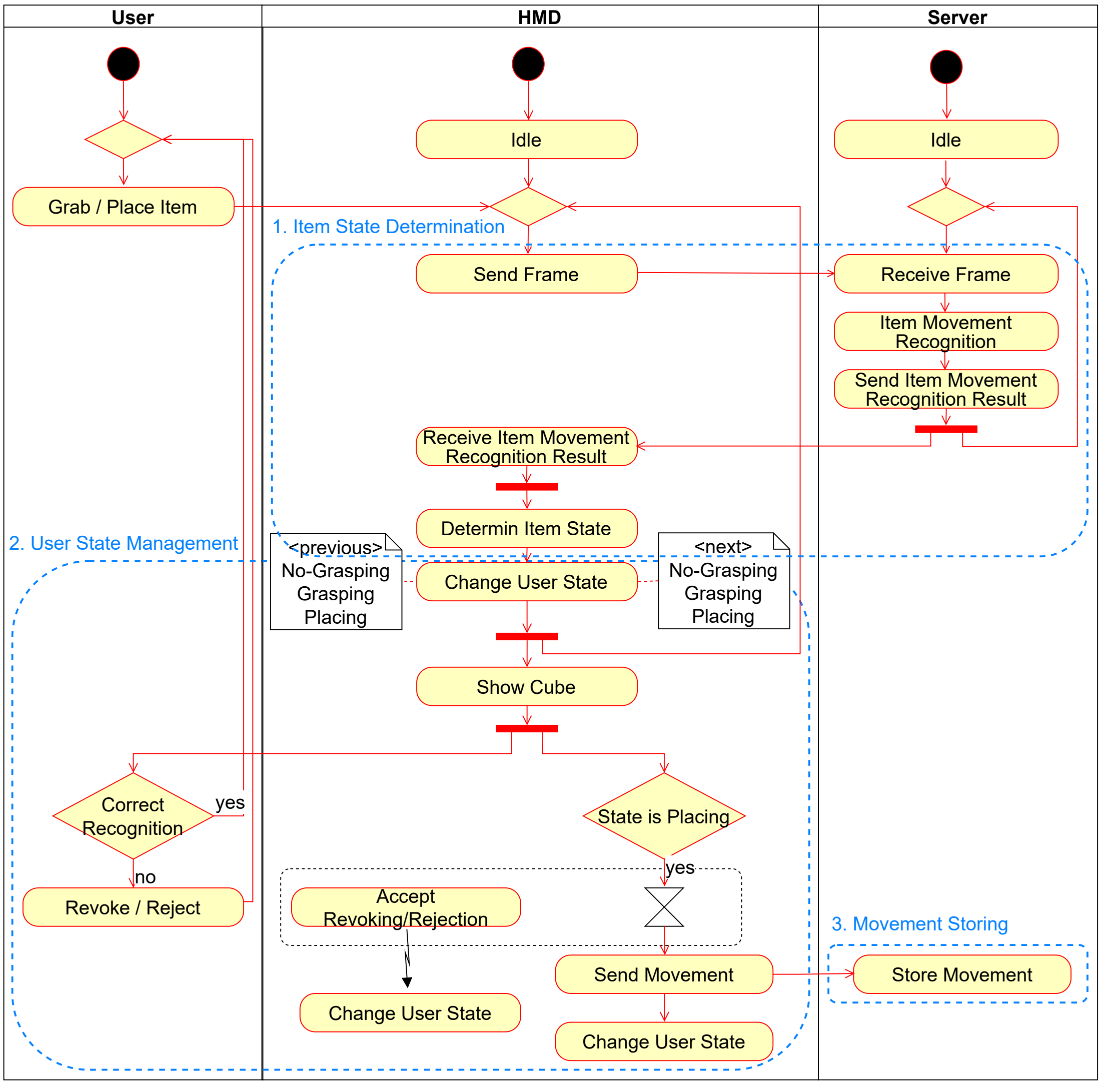


Figure 4.1: Responsible area of sections in registration.

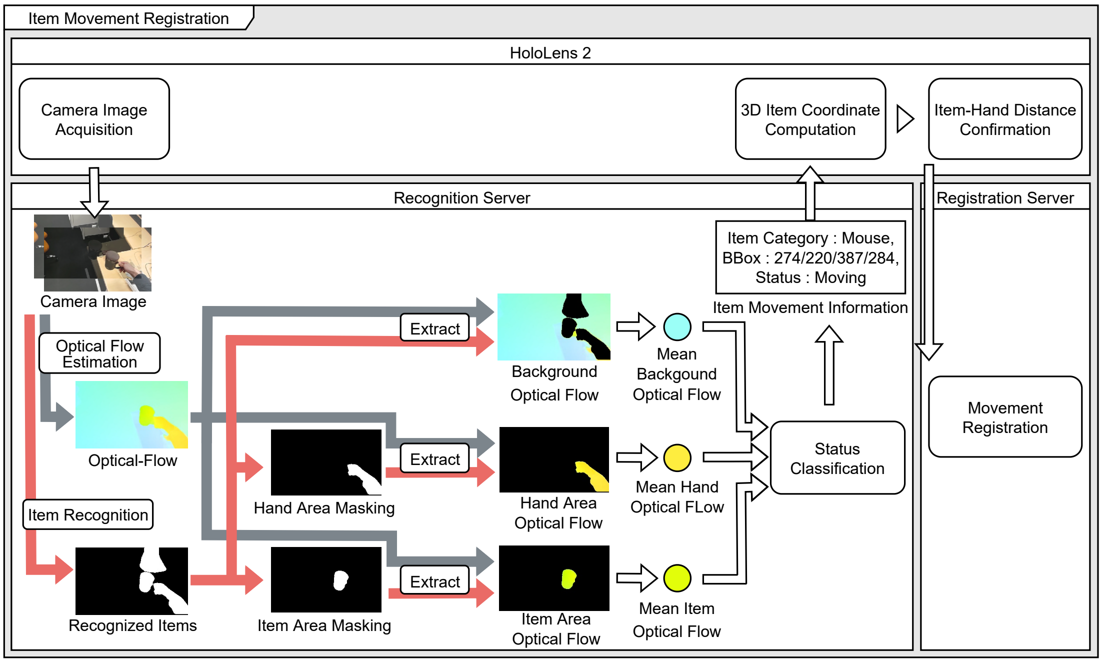


Figure 4.2: Item movement registration.

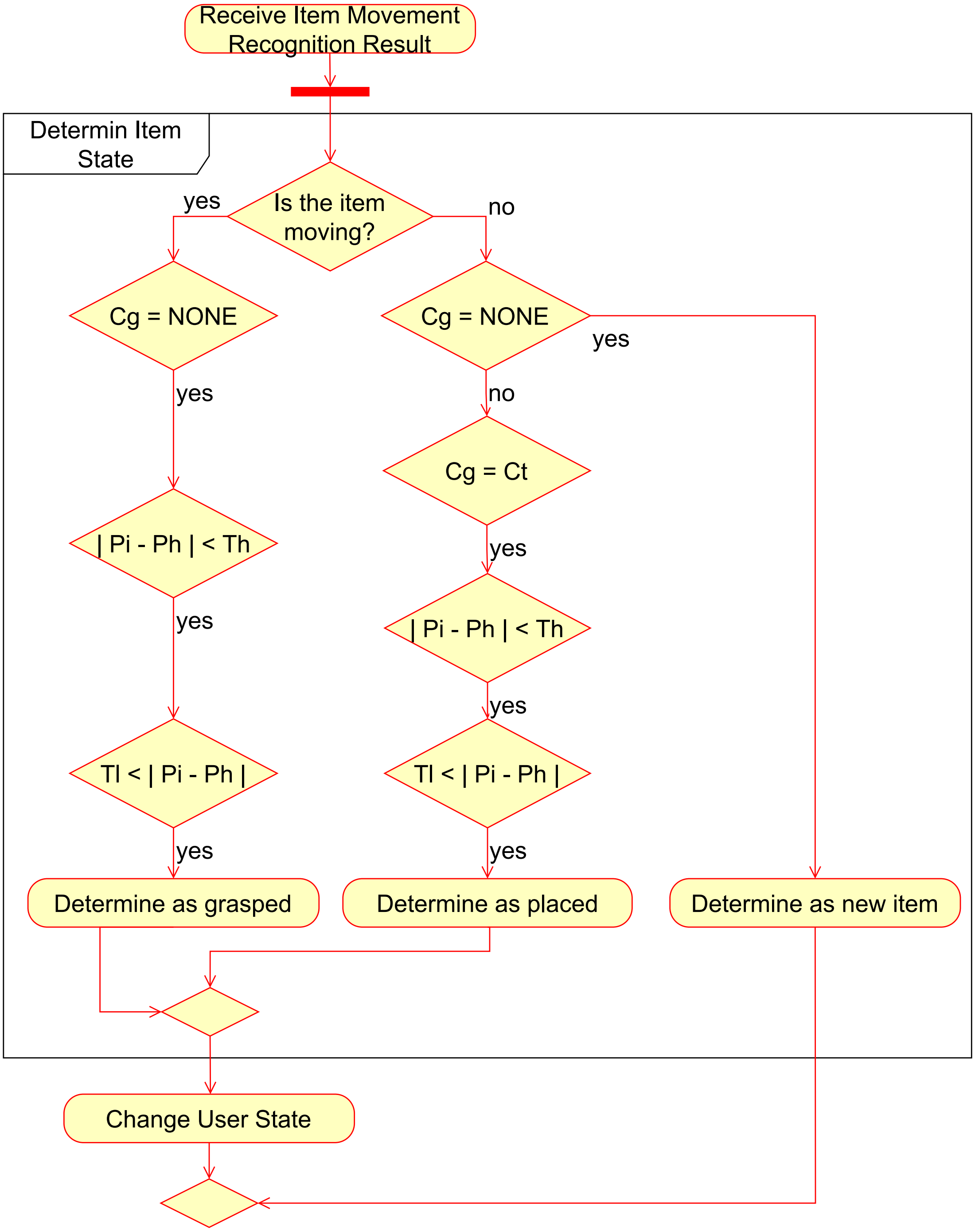


Figure 4.3: Determination of item state.

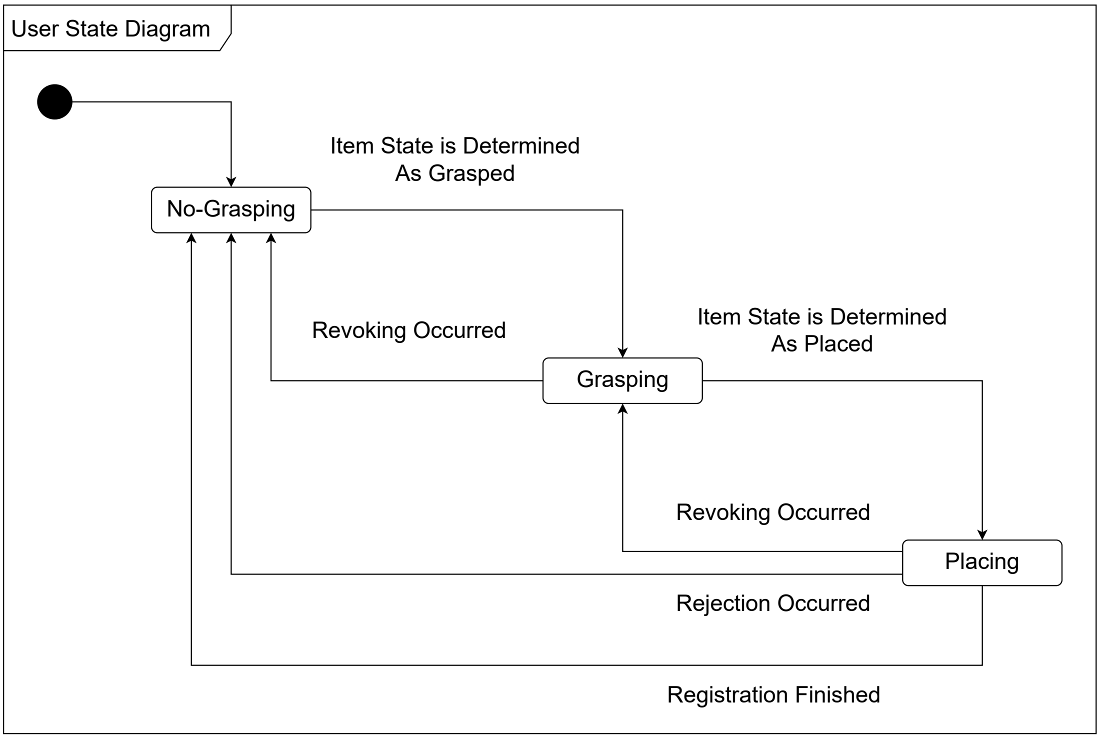


Figure 4.4: User state diagram.

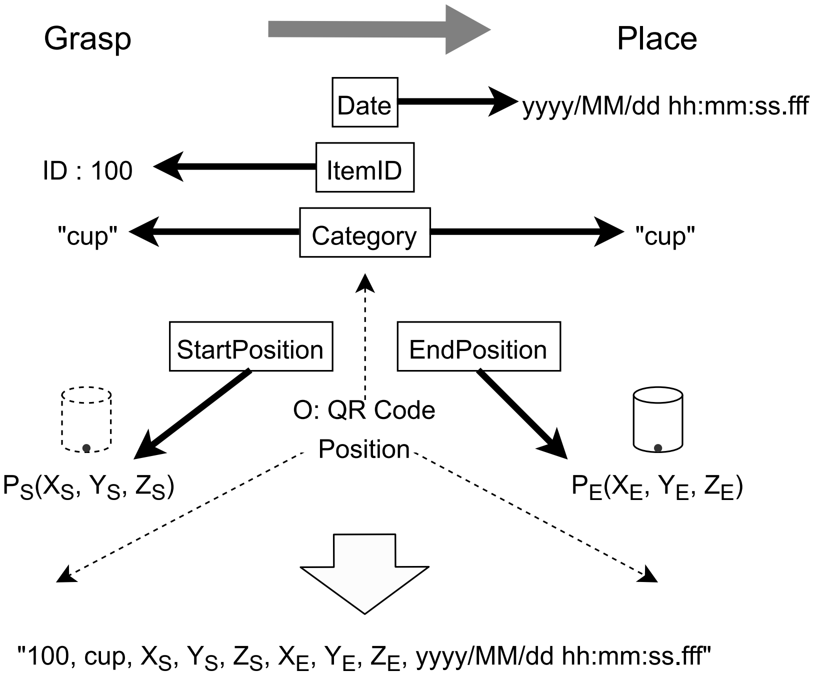


Figure 4.5: Movement storing.

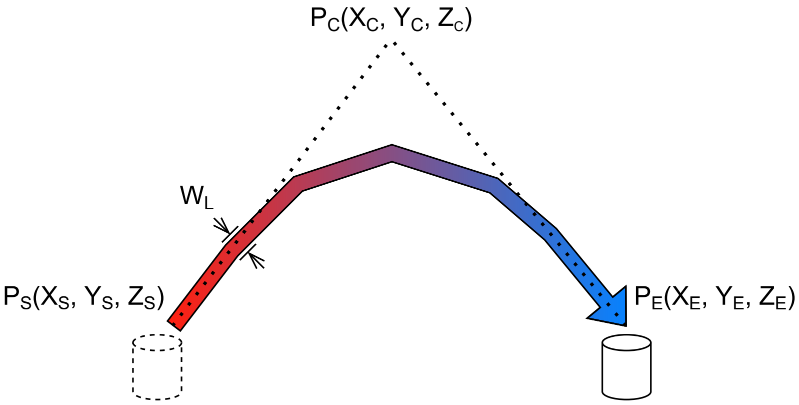


Figure 4.6: Determination of item state.

5 Parameter Optimization Experiment for Movement Recognition

5.1 Objective

We have set the parameters in an empirical manner for the optical-flow-based movement recognition system. To confirm the impact of parameters on the system and to optimize the recognition, we perform parameter optimization in the registration experiment.

5.2 Target Parameters

We have the following four target parameters:

- The number of iterations of RAFT in an optical flow estimation.
- The threshold of confidence in YOLACT++'s item recognition.
- The threshold of optical flow difference between item-background difference and item-hand difference.
- The size of the image used in RAFT and YOLACT++'s processing.

The optical flow estimation depends on the number of iterations and the size of the input image. The higher these parameters, the more accurate the optical flow is generated, but the processing time also increases. Therefore, we consider the optical flow generation of the RAFT parameters' to be the most critical parameter. Also, optical flow generation has the most significant impact on the processing speed of the entire system. Therefore, there is a tradeoff between improved movement recognition accuracy due to high-precision optical flow and the risk of missed movements due to increased processing time. From the above, the optimization of these parameters is considered to contribute significantly to increasing the system's availability, and we built a system for the optimization.

5.3 Procedure

The optimization system uses actual registration and the grid search method to perform the optimization. Within the simulation process, the movement recognition server acquires frames from the recorded video clip prepared for evaluation and outputs recognition results of grasping and moving items after performing the same processing as during actual use. The recognition results are compared with the correct data to evaluate the parameters and search for the optimal parameter set. We have the system perform a binary classification of whether a video contains the movement of an item of a certain category, and the F1 score for each set of parameters is used as the evaluation metric. The classification is positive when an item of a certain category has been moved, and negative otherwise, and if the movement of the wrong category is recognized, it is treated as a false negative.

The interval between frame acquisitions from the video clip depends on the processing speed of the system, which is subject to change with the parameters. Therefore, the total time of the video clip will be taken to simulate, since frame acquisition must be done at the speed of processing for each parameter set. The processing time of the HoloLens 2 application must also be considered. The response time of HoloLens 2 is different when the recognition server does not recognize an item, when it recognizes a moving item, and when it recognizes a stationary item, respectively. The average response times measured in advance were 58 [ms], 46 [ms], and 62 [ms], respectively, and these were added to the processing time of the simulation to get closer to the actual processing time.

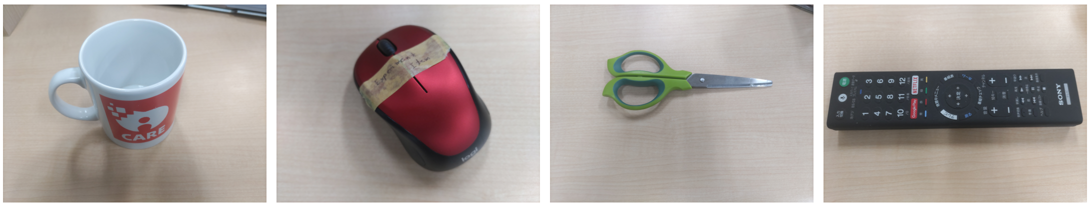
5.4 Dataset

The video clips used for optimization were recorded at the start of grasping and moving of an item at three different speeds for the four categories of items (cup, mouse, scissors, and remote) as shown in Fig. 5.1 in the COCO dataset used to train YOLACT++. The participants were asked to grasp one of these items at the following three speeds and it is performed as Fig. 5.2. First, as shown in Fig. 5.2a, the participants wait until the system recognizes the QR code and items. After the recognition, as shown in Fig. 5.2b, participants grasp the items. Finally, as shown in Fig. 5.2c, the item is brought out of the participants' sight and handed to the experimenter.

- Grab the item at a slow speed so that you will not break it.
- Grab the item at a normal speed.
- Grab the item at a fast speed as if you were in a hurry.

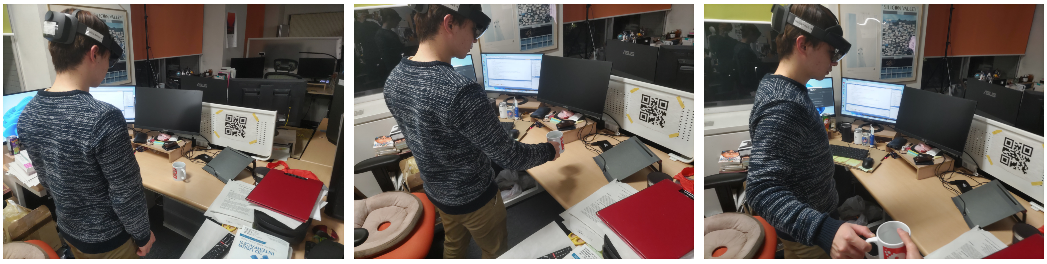
We set the lower and upper limits of the grasp movement speeds covered by the system to 1 and 3, respectively. In this manner, we do not specify a specific speed for the grasping movement, but we use the data of the speeds that the subject considers from the definition of the respective speeds. To evaluate the misrecognition of grasping, we also recorded the condition in which the participants only waved their hands over the item without grasping it at these three different speeds. Sampling was conducted for a total of five participants recruited from the laboratory of the first author, for a total sample size of 120 (4 items, 3 speeds, grab/non-grab, and 5 participants), for a total playback time of 185 seconds. Note that the order of video collection was not counterbalanced. This was decided to avoid confusing subjects and sampling unnatural movements by making the order of speed and categories irregular.

Because the system uses images from a camera mounted on the front part of the HMD, if the user grasps the item outside of the camera's angle of view, the system cannot recognize it. To ensure that the expected behavior was captured in the sample video, we required the participants to initiate the grasping of the item within the angle of view of the camera. Unlike the original use, the angle



(a) cup (b) mouse (c) scissors (d) remote

Figure 5.1: Item categories.



(a) Stare at the target item. (Before Grasping) (b) Grasp the target item (c) Hand the item to the experimenter. (After Grasping)

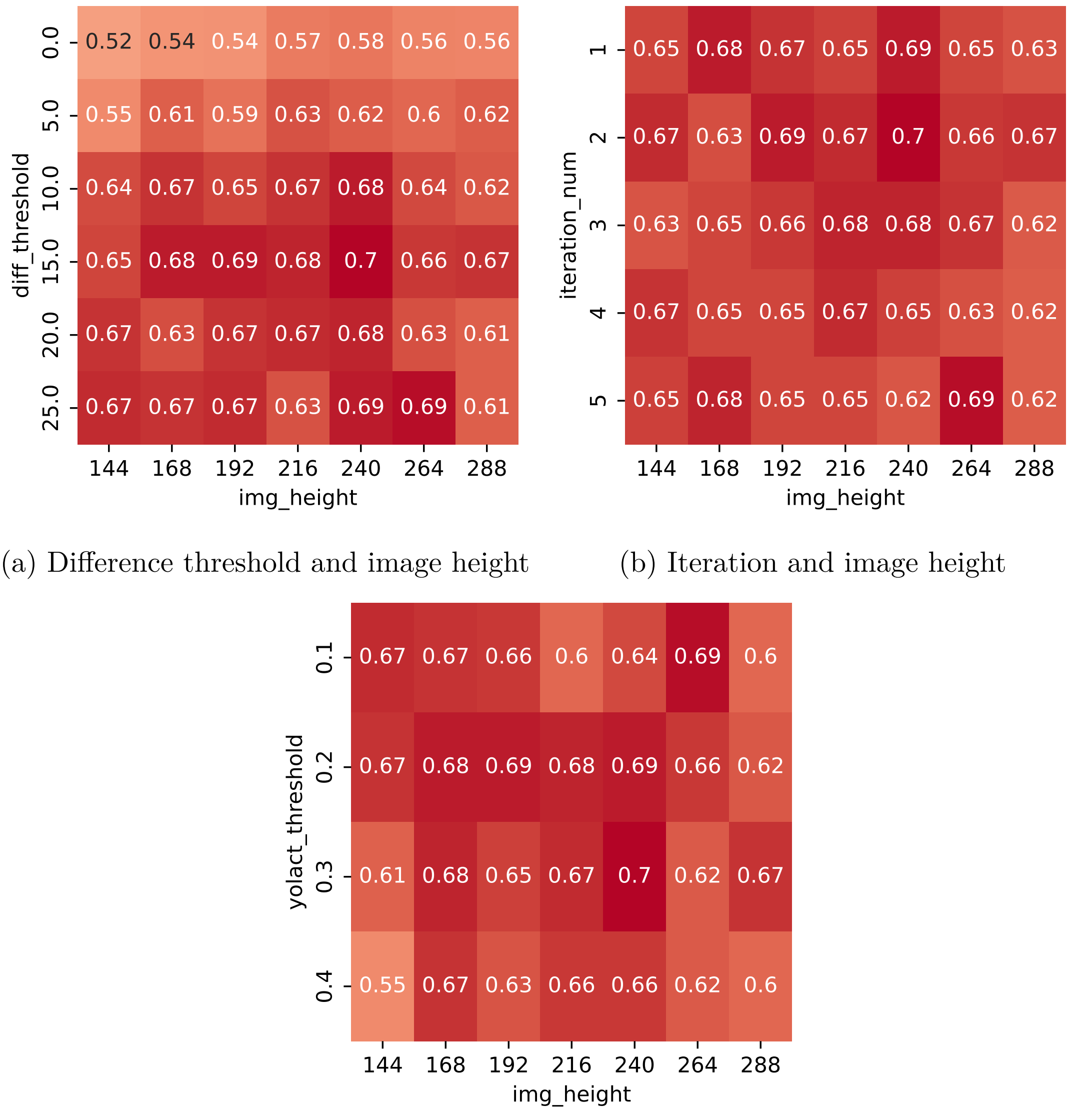
Figure 5.2: Video Collection

of view of the camera was indicated by presenting the participants with a red border as shown in Fig.5.4a.

5.5 Result

The experiment resulted in the highest F1 score of 0.7. The optimization results are distributed as shown in Fig. 5.3. The effect of optimization can be observed for all parameters, with a 0.28 improvement in F1 score from 0.42 for the empirical parameter set. The processing time per frame after optimization was 114 [ms], a reduction of 163 [ms] from 277 [ms] before optimization.

The most significant difference between the optimized parameter set and the empirical parameter set was the extremely low number of iterations in the optical flow estimation. In the empirical parameter set, this was set to 25, while in the optimal parameter set, this value was 2. As shown in Fig. 5.4, the accuracy of the optical flow obtained with one iteration or two iterations and higher iterations differs considerably. Therefore, we set a higher number of iterations to obtain a more accurate optical flow. However, since our movement recognition algorithm uses the average of the optical flow in the segmented region of the image, the required accuracy of the optical flow is low, and it is desirable to maintain a high frame rate by setting the number of iterations to a small value.

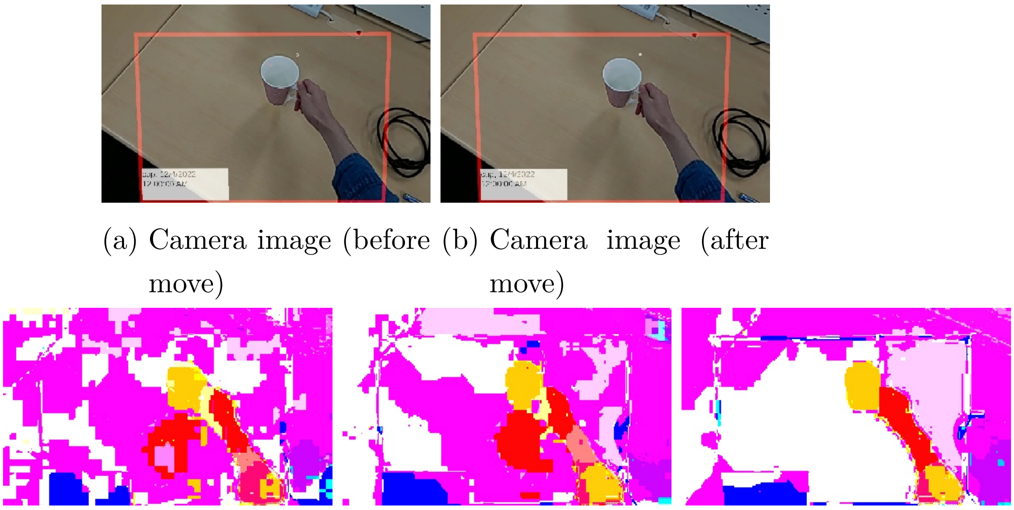


(a) Difference threshold and image height

(b) Iteration and image height

(c) YOLACT threshold and image height

Figure 5.3: Visualization of optimization result.



(a) Camera image (before move)

(b) Camera image (after move)

(c) Optical flow (1 iteration)

(d) Optical flow (2 iterations)

(e) Optical flow (10 iterations)

Figure 5.4: Number of iterations and optical flows.

6 Search Assistance Experiment

6.1 Objective

To confirm the effectiveness of our system, we compared the user's experience and time of finding items with and without the system.

6.2 Procedure

6.2.1 Participants

16 volunteers (14 males and two females in their twenties and thirties) recruited from the laboratory of the first author participated in the study. All of them had a prior experience with AR applications that use AR glasses. The participants spend a considerable amount of time in the laboratory where the experiment was conducted. This experiment was conducted in line with the Helsinki Declaration and approved by the Ethical Board of our institution.

6.2.2 Survey on item movement

To reproduce the item search task as realistically as possible, a questionnaire survey was conducted in advance with the students who work in the laboratory where the experiment took place. In this questionnaire, the participants were asked to select the locations where misplacement, such as leaving items behind, is likely to occur for the previously listed categories of items to be searched for. They had to choose those locations from candidate locations such as desks and shelves shown in Fig. 6.1. The size of the search area is approximately 240 $[m^2]$. The misplacement scenario obtained from this questionnaire was used for the item search task in this experiment.

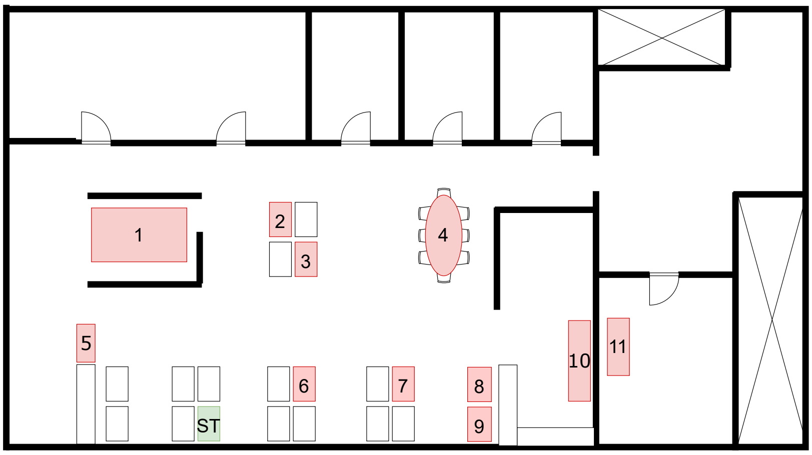


Figure 6.1: Candidate places of misplacement. ST means a place to conduct positioning.

The starting and ending points of item movement that were obtained from the questionnaire were distributed as shown in Fig. 6.2, indicating that the following characteristics are evident for each category.

- Cup: Both the starting and ending points of movement are concentrated in location 4 (snack tables).
- Scissors: Movement start points are concentrated in location 10 (stationery shelves), and movement endpoints are concentrated in location 4.
- Mouse: The start and end points of movement are close to each other in many cases and far apart in a few cases.
- Remote controller: The movement start point is concentrated at location 1 (relaxing corner), and the movement endpoint is distributed in locations 2 (office table), 4 and 5 (tool table).

6.2.3 Search items

The same categories as in the optimization experiment were used: cups, scissors, mice, and remote controller. These items are actually used daily in the laboratory.

6.2.4 Task and procedure

The participants searched for items as Fig. 6.3. We measured the time it took them to find the items. After the experiment, a questionnaire survey was conducted. In the questionnaire, the degree of anxiety felt during the item search was surveyed.

The test cases for the task were created by dividing the total of 64 movements randomly obtained from the preliminary questionnaire into 16 groups of 4. Two groups of movements were randomly assigned per participant. One group was searched with the system and the other group was searched without the system. For each search task, the experimenter placed one item and asked the participant to search for it as Fig. 6.3a and touch it as Fig. 6.3b. This was repeated 16 times for each participant. During the placement of the item, participants were blindfolded and wore headphones so that they remained clueless about the location of the item.

When not using the assistance system, we showed the participant a picture of the item to be searched for. Then we ask the participant to start search for the item without wearing the AR HMD by the experimenter's cue. We measured the time from the moment the participant started moving to the moment the participant found the item and touched it. When using the assistance system, we had the participant perform Air Tap by the experimenter's cue, and measured the time it took the participant to find and touch the item using the search assistance function. In cases when the Air Tap was not recognized after the cue due to the surrounding environment or other reasons, the measurement was started after confirming that the Air Tap was recognized and the movement was visualized.

6.3 Results

6.3.1 Search Time

The distribution of the time taken to search is shown in Fig. 6.4. A two-way ANOVA was performed to analyze the effect of system presence/absence and item category on search time. A two-way ANOVA revealed that there was not a statistically significant interaction between the effects of system presence/absence and item category ($F(3, 45) = 1.053, p = 0.378, \eta_p^2 = 0.066$). Simple main effects analysis showed that system presence/absence did have a statistically significant effect on search time ($F(1, 15) = 8.180, p = 0.012, \eta_p^2 = 0.353$). Simple main effects analysis showed that item category did not have a statistically significant effect on search time ($F(3, 45) = 1.110, p = 0.355, \eta_p^2 = 0.069$).

6.3.2 Questionnaire

16 participants answered a questionnaire containing four questions below at the end of each of two series of item-finding tasks with and without the system.

- Q1: I felt I performed the search efficiently.
- Q2: I felt anxious during the finding.
- Q3: I felt comfortable during the finding.
- Q4: I felt tired during the finding.

A Wilcoxon signed-ranks test indicated that the rank of the efficiency they felt during search tasks with our system was statistically significantly higher than the one without our system ($W = 0.0, Z = -3.552, p < 0.001, r = -0.8880$). The rank of the anxiety they felt during search tasks with our system was statistically significantly lower than the one without our system ($W = 2.0, Z = -3.203, p < 0.005, r = -0.801$). The rank of the comfort they felt during search tasks with our system was statistically significantly higher than the one without our system ($W = 0.0, Z = -3.557, p < 0.001, r = -0.8892$). The rank of the tiredness they felt during search tasks with our system was statistically significantly lower than the one without our system ($W = 0.0, Z = -3.443, p < 0.001, r = -0.8607$). The results of the questionnaire are shown in Fig. 6.5 and indicate that the system significantly reduces the burden on the user in the task in all question items.

6.3.3 Observation

The participants' behavior during the experiment revealed that their daily line of movement, as well as the simple travel distance of the item, affected the search time. The endpoints of Scissors' movements were concentrated in location 4 (snack table), and when the search started at location 10, 4 were visible from the starting point, and most participants moved in a straight line to location 4 after taking a glance at locations 8 and 9 (office tables). Thus, the participants' movement routes were generally similar with and without the assistance system. In contrast, for the remote controller, they behaved differently with and without the system. There was a split from starting point 1 to 2 or 5. A time loss occurred if they made the wrong choice between the two places. Also, for those without the system, overlooking an item at a location where the item was actually present caused them not to return to the same location for an extended period of time.

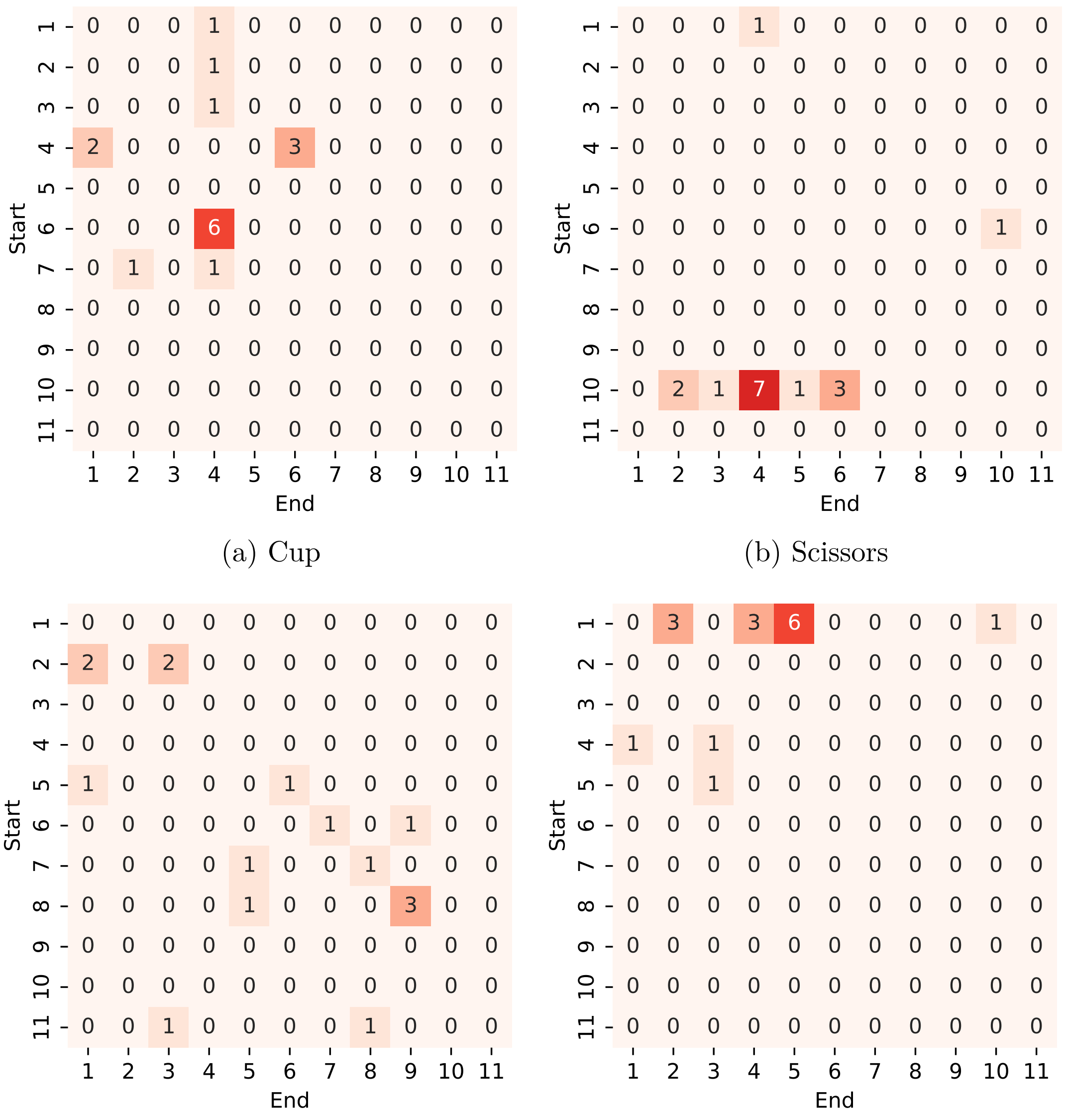
6.3.4 Feedback

In the post-experiment questionnaire, the following opinions were obtained from the participants regarding the usability of the system. They are on the characteristics of the hardware device and the implementation of the system.

Seven out of the 16 participants commented on the recognition accuracy of the Air Tap and their familiarity with its operation (participants A, B, C, E, H, I, and N). Participant I mentioned: *"I wasn't used to the air pinch and sometimes it didn't respond, but once I got used to it, I didn't have too much trouble."*

When participants commented on the highlighting of the end of the move, there were some negative opinions on the effect on guidance and others: *"It was easy to see that the location of the item is marked in green."* (participant B); *"Since the lines alone were enough for tracking, I did not feel a strong need for highlighting."* (participant C); *"When arriving at a target location, it was difficult to visually see the surrounding area due to the highlight. It would be better to adjust the transparency when approaching the highlighted area or use a border display."* (participant H).

While positive opinions were expressed about the sphere selection method itself, there were cases where spheres were buried in walls or where several large spheres were adjacent to each other, making it difficult to use the system: *"It was easy to access the movement by touching the sphere."* (participant J); *"Sometimes the ball was buried in the shelf and I couldn't touch it and highlight the destination."* (participant H); *"When the large white balls overlapped, I could recognize up to* *two of them, but it was difficult to see the third one."* (participant I).



	1	2	3	4	5	6	7	8	9	10	11
1	0	0	0	1	0	0	0	0	0	0	0
2	0	0	0	1	0	0	0	0	0	0	0
3	0	0	0	1	0	0	0	0	0	0	0
4	2	0	0	0	0	3	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	6	0	0	0	0	0	0	0
7	0	1	0	1	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0

(a) Cup

	1	2	3	4	5	6	7	8	9	10	11
1	0	0	0	1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	1	0
7	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0
10	0	2	1	7	1	3	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0

(b) Scissors

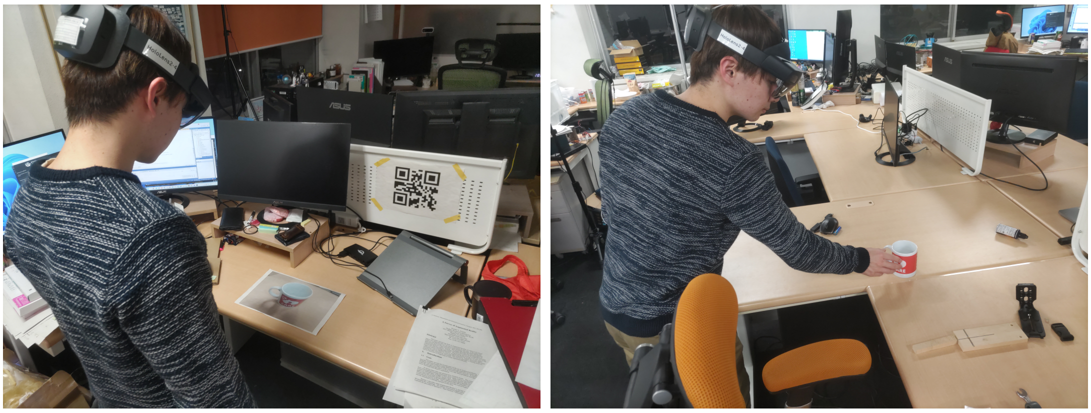
	1	2	3	4	5	6	7	8	9	10	11
1	0	0	0	0	0	0	0	0	0	0	0
2	2	0	2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0
5	1	0	0	0	0	1	0	0	0	0	0
6	0	0	0	0	0	0	1	0	1	0	0
7	0	0	0	0	1	0	0	1	0	0	0
8	0	0	0	0	1	0	0	0	3	0	0
9	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0
11	0	0	1	0	0	0	0	1	0	0	0

(c) Mouse

	1	2	3	4	5	6	7	8	9	10	11
1	0	3	0	3	6	0	0	0	0	1	0
2	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0
4	1	0	1	0	0	0	0	0	0	0	0
5	0	0	1	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0

(d) Remote controller

Figure 6.2: Distribution of survey results on plausible misplacement items.



(a) Start Search

(b) Stop Search

Figure 6.3: search Experiment

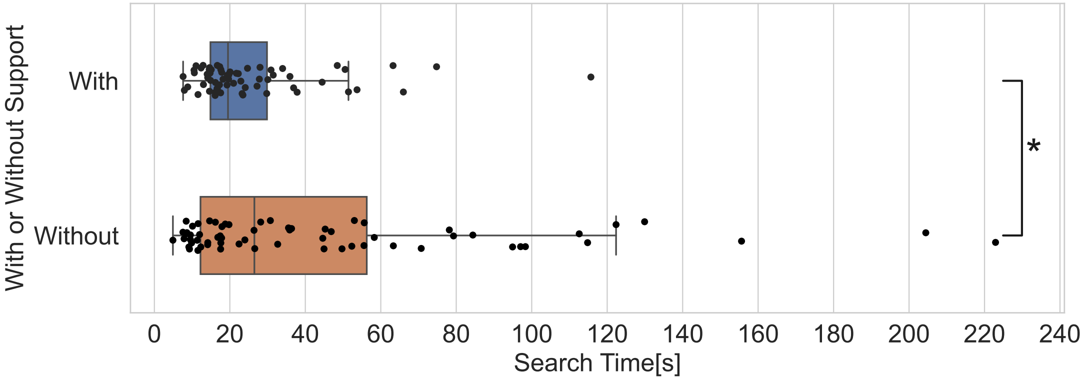


Figure 6.4: Result of search time.

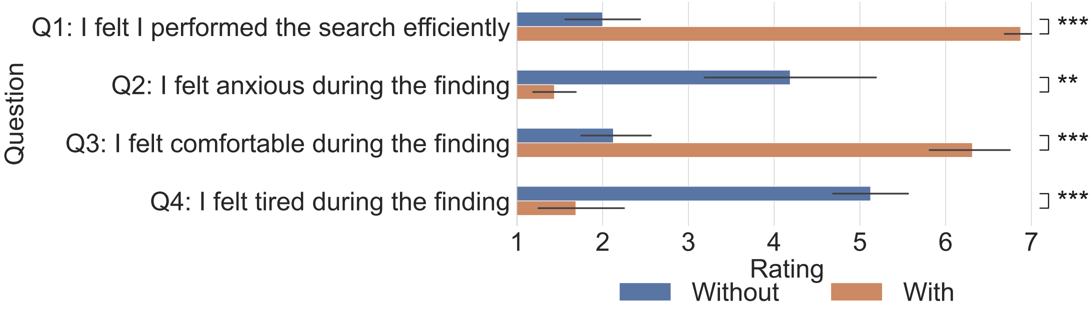


Figure 6.5: Result of questionnaire.

7 Discussion

7.1 Usefulness of The System

7.1.1 Search function

In a shared environment such as an office where a large number of items are moved around, image-based retrieval has its limitations. If an image is missed because the user does not remember the item's appearance or for some other reason, the user has to go around the image list several times to find the image. Since our system uses location for search, it is possible to narrow down the vast amount of movement to a very small number of options, as long as the user knows where the item has existed in the past. Therefore, this method is effective in reducing the risk of missing item movement.

In addition, systems that use pre-registration to manage items require time and monetary costs for pre-registration solutions such as RFID tags to manage items, and the cost depends on the scale of the number of items to be managed. Our system is expected to reduce this preparation cost because it automatically recognizes items.

7.1.2 Guidance

Conventional guidance techniques using images and short videos may not be able to pinpoint the locations because the user may not be familiar with the appearance of the area of concern. The visualization of our system is similar to the arrow-and-distance visualization of Funk et al. in the aspect of location presentation [5]. However, this method does not present a pinpoint location to assist in locating items at close range, instead, it presents an image of the item's surroundings. Their method assumes that the presentation of images will provide a contextual clue, but it is questionable how effective this will be if the user has no knowledge of the location.

Our method presents pinpoint locations and does not require the user to have contextual knowledge about the destination location. In this respect, we believe that our proposed method is more effective in assisting users to find things in a wider range of locations and users. As Osmers' study shows [30], AR display methods can have a combined effect of amplification, compensation, and complementation. So, future work combining other visualization methods with our current method will be very meaningful.

7.1.3 Privacy

In addition to our system, other systems, such as "GO-Finder" [1], perform automatic registration without requiring pre-registration of items, but existing automatic registration systems can cause various privacy issues when applied to multiple users. Our system is expected to reduce the barriers to multi-user use by providing countermeasures to the problems of other people's presence in the image and the segregation of jointly managed items and personal property.

7.1.4 Use in collaborative space

When using the system in a cooperative environment, influence on the real environment and others should be avoided. A method that physically hightlights the location of an item, such as using a real spotlight [3], is likely to disturb others in a collaborative space. In this respect, our system is more suitable for use in a cooperative environment because it can reduce the impact on others by using AR.

7.2 Limitations

7.2.1 Size of items

Since the system is designed to recognize the grasp of an item by the user's hand, the problem of severe covering of small items can naturally happen, which is also discussed in the grasp recognition study we refer [27]: small objects tend to be occluded and would be problematic considering real-world use. One potential solution is to use the object' s appearance before and after manipulation but not during manipulation with hand occlusion [27]. It would be difficult to deal with this problem with only one of the images before or after the grasp. This is because when there are multiple small items, which tend to be hidden by the hand, making it impossible to tell which item was grasped.

7.2.2 Items of same category

In the current system, when multiple items in the same category move in a short period of time from a nearby location, the user can differentiate them only by the small differences in their starting points and time. If there are only two or three items of movement that can be visualized under such conditions, the user may be able to find them by visiting all of them, but the significance of this system may diminish as the number increases.

However, as we wrote in the beginning, we target only shared items. Considering the cooperative environment in which the system is used, if multiple items of the same category exist in the same location, it is highly likely that this is the original place where they are managed. Therefore, the best way to deal with this problem is not to identify what items have been taken out of the original place, but to provide a way to know if they have been returned. This can be accomplished by combining the movement of the same item into one. Since multiple items are never in the exact same location at the same time, it is possible to treat the location and time of their existence as the identifier of the item. If the end point of one movement is close enough to the start point of another movement and there are no other movements in between, it would be possible to treat them as movements of the same item.

8 Conclusion

In this paper, we have developed a multifunctional, auto-registering, item management system for use in a collaborative environment. The proposed a system provides automatic registration, search function, guidance function, and privacy protection with the following key ideas.

- Automatic registration of item movement by grasping without the necessity of pre-registration.
- Search function for item movement using location as a query key.
- Visualization of the location before and after the item movement.
- Elimination of image from registered data to preserve privacy.
- A UI for rejection of registration.

The parameter optimization conducted in the automatic registration experiment led to the discovery of optimal values that are difficult to reach in an empirical manner. The item search experiment showed that the search and guidance functions of the proposed system are effective as an assistance system for item search, both in terms of search time and user experience. Our future work includes improving grasp-and-move detection and a longitudinal evaluation with multiple users in a shared space.

Acknowledgements

I am happy to have learned about the VR and AR fields over the past two years and to have built an AR system based on my original idea. I am also grateful to all the people who constantly supported my research. I appreciate Prof. Kiyokawa, Assoc. Prof. Uchiyama, Asst. Prof. Isoyama, and Asst. Prof. Perusquía-Hernández for their support from the idea of my research to the submission of my thesis. I would also like to thank Prof. Yasumoto for co-supervising my research and checking the direction of my research. Completing my research with my own idea and my own system gave me a lot of confidence. Ph.D. students Mr. Hagimori and Mr. Nakano consulted with me on various aspects such as experimental techniques and analysis methods and gave me many valuable opinions as an experiment's participants. Thanks to their valuable time, I was able to conduct very good research. I am grateful to all the members of this laboratory for making my life as a student in the master's program a healthy one. Mainly, I have very good memories of working with Mr. Kubota, Mr. Matsuo, and Mr. Yokoro for 6 months on the same research project and winning a prize in a contest. Mr. Miyawaki, Mr. Otono, Mr. Miyazaki, and Mr. Sasaki consulted with me on research policies and technical aspects and gave me various valuable opinions. Mr.Aoki, Mr.Fujisawa, and Ms.Otsuka supported my student life at various times. I have too many memories to list here, and my student life would have been completely different without them. I thank everyone who has been involved in my life as a student over the past two years.

Bibliography

[1] 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 *Proceedings of the 26th International Conference on Intelligent User Interfaces*, IUI '21, pp. 139–149, 2021.

[2] 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 *Proceedings of the Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI EA '22, pp. 1–7, 2022.

[3] T. Nakada, H. Kanai, and S. Kunifuji, "A support system for finding lost objects using spotlight," in *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services*, MobileHCI '05, pp. 321–322, 2005.

[4] K. Chen, Y. Huang, Y. Chen, H. Zhong, L. Lin, L. Wang, and K. Wu, "Lisee: A headphone that provides all-day assistance for blind and low-vision users to reach surrounding objects," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 3, pp. 1–30, 2022.

[5] M. Funk, R. Boldt, B. Pfleging, M. Pfeiffer, N. Henze, and A. Schmidt, "Representing indoor location of objects on wearable computers with head-mounted displays," in *Proceedings of the 5th Augmented Human International Conference*, AH '14, 2014.

[6] F. M. Li, D. L. Chen, M. Fan, and K. N. Truong, "Fmt: A wearable camera-based object tracking memory aid for older adults," *Proceedings of the ACM*

on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 3, no. 3, pp. 1–30, 2019.

[7] M. Funk, A. Schmidt, and L. E. Holmquist, “Antonius: A mobile search engine for the physical world,” in *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication*, UbiComp ’13 Adjunct, pp. 179–182, 2013.

[8] Y. Á. López, J. Franssen, G. Á. Narciandi, J. Pagnozzi, I. G.-P. Arrillaga, and F. L.-H. Andrés, “RFID technology for management and tracking: e-health applications,” *Sensors*, vol. 18, no. 8: 2663, pp. 1–17, 2018.

[9] G. Borriello, W. Brunette, M. Hall, C. Hartung, and C. Tangney, “Reminding about tagged objects using passive rfids,” in *Proceedings of the 6th International Conference on Ubiquitous Computing*, UbiComp ’04, pp. 36–53, 2004.

[10] X. Liu, M. D. Corner, and P. Shenoy, “Ferret: Rfid localization for pervasive multimedia,” in *Proceedings of the 8th International Conference on Ubiquitous Computing*, UbiComp ’06, pp. 422–440, 2006.

[11] M. Elsayeh, M. Haroon, B. Tawfik, and A. Fahmy, “Rfid-based indoors localization of tag-less objects,” in *Proceedings of the 5th Cairo International Biomedical Engineering Conference*, CIBEC ’10, pp. 61–65, 2010.

[12] M. Tanbo, R. Nojiri, Y. Kawakita, and H. Ichikawa, “Active rfid attached object clustering method with new evaluation criterion for finding lost objects,” *Mobile Information Systems*, vol. 2017, pp. 1–12, 2017.

[13] P. Wilson, D. Prashanth, and H. Aghajan, “Utilizing rfid signaling scheme for localization of stationary objects and speed estimation of mobile objects,” in *Proceedings of the 2007 IEEE international conference on RFID*, IEEE RFID ’07, pp. 94–99, 2007.

[14] J. A. Kientz, S. N. Patel, A. Z. Tyebkhan, B. Gane, J. Wiley, and G. D. Abowd, “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, pp. 103–110, 2006.

[15] L. Pei, R. Chen, J. Liu, T. Tenhunen, H. Kuusniemi, and Y. Chen, "Inquiry-based bluetooth indoor positioning via rssi probability distributions," in *Proceedings of the 2nd International Conference on Advances in Satellite and Space Communications*, SPACOMM '10, pp. 151–156, 2010.

[16] D. Schwarz, M. Schwarz, J. Stückler, and S. Behnke, "Cosero, find my keys! object localization and retrieval using bluetooth low energy tags," in *Proceedings of the RoboCup 2014: Robot World Cup XVIII*, pp. 195–206, 2015.

[17] A. Farasin, F. Peciarolo, M. Grangetto, E. Gianaria, and P. Garza, "Real-time object detection and tracking in mixed reality using microsoft hololens," in *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, vol. 4 of *VISIGRAPP '20*, pp. 165–172, 2020.

[18] T. Maekawa, Y. Yanagisawa, Y. Kishino, K. Ishiguro, K. Kamei, Y. Sakurai, and T. Okadome, "Object-based activity recognition with heterogeneous sensors on wrist," in *Proceedings of the 8th International Conference on Pervasive Computing*, Pervasive'10, pp. 246–264, 2010.

[19] E. Wu, Y. Yuan, H.-S. Yeo, A. Quigley, H. Koike, and K. M. Kitani, "Back-hand-pose: 3d hand pose estimation for a wrist-worn camera via dorsum deformation network," in *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, UIST '20, pp. 1147–1160, 2020.

[20] T. Ueoka, T. Kawamura, Y. Kono, and M. Kidode, "I' m here!: A wearable object remembrance support system," in *Proceedings of the 5th Human-Computer Interaction with Mobile Devices and Services*, Mobile HCI '03, pp. 422–427, 2003.

[21] R. Hoyle, R. Templeman, S. Armes, D. Anthony, D. Crandall, and A. Kapadia, "Privacy behaviors of lifeloggers using wearable cameras," in *Proceedings*

of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '14, pp. 571–582, 2014.

[22] T. Starner, S. Mann, B. Rhodes, J. rey Levine, J. Healey, D. Kirsch, R. W. Picard, and A. Pentland, "Augmented reality through wearable computing," *Presence: Teleoperators & Virtual Environments*, vol. 6, pp. 386–398, 1997.

[23] B. J. Rhodes, "The wearable remembrance agent: A system for augmented memory," *Personal Technologies*, vol. 1, no. 4, pp. 218–224, 1997.

[24] R. Suomela and J. Lehikoinen, "Taxonomy for visualizing location-based information," *Virtual Reality*, vol. 8, no. 2, pp. 71–82, 2004.

[25] R. P. Darken and J. L. Sibert, "Wayfinding strategies and behaviors in large virtual worlds," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '96, pp. 142–149, 1996.

[26] R. T. Azuma, "A survey of augmented reality," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, pp. 355–385, 1997.

[27] T. Yagi, M. T. Hasan, and Y. Sato, "Hand-object contact prediction via motion-based pseudo-labeling and guided progressive label correction," in *Proceedings of the 32nd British Machine Vision Conference*, BMVC '21, pp. 1–14, 2021.

[28] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLACT++ better real-time instance segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 2, pp. 1108–1121, 2022.

[29] Z. Teed and J. Deng, "Raft: Recurrent all-pairs field transforms for optical flow," in *Proceedings of the 16th European Conference on Computer Vision*, ECCV '20, pp. 402–419, 2020.

[30] N. Osmers and M. Prilla, "Getting out of out of sight: Evaluation of ar mechanisms for awareness and orientation support in occluded multi-room settings," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pp. 1–11, 2020.

Publication List

[1] Shota Matsuo, Hiroto Oshimi, Taichi Kubota, Kaito Yokoro, Naoya Isoyama, Hideaki Uchiyama, and Kiyoshi Kiyokawa , “Nagisampo,” 2021 Interverse Virtual Reality Challenge (IVRC 2021), Sep. 2021.
[2] Hiroto Oshimi, Monica Perusquía-Hernández, Naoya Isoyama, Hideaki Uchiyama, Kiyoshi Kiyokawa, “Privacy-conserving AR-based support system for finding objects in a shared space,” 2022 Multimedia, Distributed, Cooperative, and Mobile Symposium (DICOMO 2022), July 2022.
[3] Hiroto Oshimi, Monica Perusquía-Hernández, Naoya Isoyama, Hideaki Uchiyama, Kiyoshi Kiyokawa, “LocatAR: An AR Object Search Assistance System for a Shared Space,” Augmented Humans 2023, Mar. 2023.