Authors' copy. The final version has been published in Proceedings of ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT) and pr
UbiComp 2019.

FMT: A Wearable Camera-Based Object Tracking Memory Aid for Older Adults

FRANKLIN MINGZHE LI, University of Toronto, Canada
DI LAURA CHEN, University of Toronto, Canada
MINGMING FAN, University of Toronto, Canada & Rochester Institute of Technology, USA
KHAI N. TRUONG, University of Toronto, Canada

Older adults sometimes forget about whether or not they have completed routine actions and the states of objects that they have interacted with (*e.g.,* the kitchen stove is on or off). In this work, we explore whether video clips captured from a body-worn camera every time objects of interest are found within its field of view can help older adults determine if they have completed certain actions with these objects and what their states are. We designed FMT ("Fiducial Marker Tracker")—a real-time capture and access application that opportunistically captures video clips of objects the user interacts with. To do this, the user places fiducial markers close to objects which would be captured when the marker enters the user's body-worn camera's field of view. We examine and discuss what objects this system would be best suited to track, and the usefulness and usability of this approach. FMT successfully captured direct interactions with an object at an average rate of 75.6% across all participants (SD = 9.9%). Our results also reveal how, what, and why users would use such a system for help.

CCS Concepts: • **Human-centered computing → Empirical studies in ubiquitous and mobile computing;**

KEYWORDS

Memory aid, older adults, wearable camera, object tracking, technology probe

ACM Reference format:

Franklin Mingzhe Li, Di Laura Chen, Mingming Fan, Khai N. Truong. 2019. FMT: A Wearable Camera-Based Object Tracking Memory Aid for Older Adults. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3, Article X (September 2019), XX pages. https://doi.org/10.1145/3351253

1 INTRODUCTION

The global population is ageing rapidly. In 2017, there was an estimated 962 million older adults aged 60 or over in the world. The total population of older adults over 60 is estimated to reach two billion in 2050 [43]. The majority of older adults currently live in private households [18] and prefer living independently [37]. For many older adults, independent living is possible only with assistance from friends, family, and in-home services that help with their activities of daily living [10]. This is because older adults are likely to

This work is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).
Author's address: 40 St. George Street, Room 5166, Toronto, Ontario, Canada, M5S 2E4
Author's email addresses: Franklin Mingzhe Li, franklin.li@mail.utoronto.ca; Di Laura Chen, chendi@dgp.toronto.edu; Mingming Fan, mfan@cs.toronto.edu; Khai N. Truong, khai@cs.toronto.edu

https://doi.org/10.1145/3351253

have impaired mobility, multiple chronic health conditions, and social and economic limitations [10]. Memory problems are one of the most common complaints of older adults [15, 29]. Over 20% of the older adults over 60 suffer from a mental or neurological disorder (e.g., dementia, and depression) [43]. Even older adults who are currently in good health and have the functional abilities to live independently may require help with personal care or tasks due to mental decline as their age advances [18]. Examples of tasks that are essential for older adults to maintain independent living which they might need help with include recalling whether or not they have turned off the stove, and remembering to take medication. However, Cavanaugh et al. [9] showed that objects and actions repeated on a regular basis were the two largest content categories for which older adults often have memory failures.

Fig. 1. A participant with the FMT (Fiducial Marker Tracker) system worn around the neck.

Prior research has explored how manual and automated methods can be used to help older adults with memory problems. Manual approaches, such as shopping lists [4] and diaries [40], have been well evaluated and shown to help users recall important information, but require users to put effort into inputting content. Furthermore, users cannot recall past memories from paper logs that have not been written down—meaning often users must have the foresight of what they may need to recall later. Automatic approaches, such as SenseCam [17], allow users to recall retrospective memory by recording past events automatically. However, memory aids designed using traditional automated capture approaches often continuously record and store video content, which poses potential privacy concerns [1] in addition to demanding high volumes of storage space and computational power. If used for the specific reason of needing to recall whether a routine action has been completed or the states of objects that have been interacted with, such a design would mean a large amount of data has to be unnecessarily captured when the user only needs to see records that involve specific objects. Rather than continuously capturing all the time, we propose the capture of short video clips, starting when pre-selected objects come within view of the user and ending three seconds after those objects are no longer within view. A system designed this way would store video clips showing direct interactions that a user has with these objects as well as indirectly captured video clips of when the user passes by any of the marked objects. This approach can potentially reduce privacy concerns, power consumption, as well as large amounts of storage space required by a continuously capturing system. However, we need to understand whether our proposed approach is an effective way to create memory aid to help older adults remember the states of objects that they have interacted with, and whether they have completed particular actions in the past. To answer these questions, it is important to first answer initial questions such as if the system would be able to capture useful and usable video clips. Thus, before developing a robust prototype and conducting a deployment study with such a system, we conducted an exploratory study that examined the following research questions:

RQ1. Whether object-based video clips captured from a body-worn camera every time objects of interest are found within its field of view could help older adults recall the state of an object or past interactions with this object?

We used this study to also gain an understanding about:

RQ2. How older adults would use object-based video clips to recall states of objects and their past interactions with these objects?
RQ3. What older adults would use object-based video clips to recall states and interaction histories of?
RQ4. When and why older adults would use object-based video clips to recall states of objects and their interactions with these objects?

To explore these questions, we first conducted formative interviews with 12 older adults over 65 to understand what objects or routines they most easily forget and their current approaches for help. Then, we developed FMT ("Fiducial Marker Tracker") (Fig. 3), a real-time capture and access application [39] implemented on a mobile device that allowed older adults to recall past interactions with specific objects and states of objects. To use FMT, older adults must place fiducial markers close to the objects that they are interested in tracking. Next, the system automatically records video clips when the user interacts with the desired objects and the user would then be able to re-watch the videos afterwards in FMT to recall past interactions. We used FMT as a technology probe and conducted an exploratory study with 12 older adults (above 65 years of age) to evaluate the efficacy of this approach. Through our study, we learned that approximately 75.6% (SD = 9.9%) of the direct interactions with an object could be successfully captured by a wearable camera and participants preferred using FMT to keep track of objects such as the kitchen stove and the door lock the most for safety reasons among other findings.

Fig. 2. Examples of tracked objects that the participants interacted with.

2 RELATED WORK

Retrospective memory involves remembering past experiences (episodic or autobiographical memory) and learned information (semantic memory), which is essential for older adults to maintain functional independence [30]. Unfortunately, older adults often face difficulty with retrospective memory [15, 29].

Specifically, recalling “objects” (i.e., remembering a particular physical object that will be needed later) and “routines” (i.e., actions repeated on a regular basis) are among the top three types of memory failures that older adults reported experiencing, and are the two most common types of information for which older adults used memory aids [9].

Using external memory aids to help people compensate for their memory deficits is one of the most effective methods of memory rehabilitation [11, 32]. Previous research has investigated the use of diaries [4, 40], audio contents [12, 13] and image contents (e.g., [5, 8, 17, 22]) to help people recall past events. For example, Verbrugge [40] evaluated and showed the usefulness of diaries as a retrospective memory aid. However, Brewer [6, 7] noted that the human autobiographical memory (about past experiences) is typically rich in visual imagery. Thus, many systems have examined how to use images and videos to aid recall. For example, Jebara et al. [19] developed DyPERS, a video replay system with a body-worn camera which stores the user’s visual and auditory scenes. They conducted a comparative study with diaries and no aids to recall specific information of various objects, and showed that participants who used DyPERS had a higher accuracy in recalling particular objects. However, DyPERS required users to manually control when they wanted to capture the visual and audio scenes, and Niforatos et al. [31] showed that manual picture capture might lead to the encoding of memories of lower quality.

Automatic approaches have the advantages of reducing the workload of manual logging and support the capture of images without the user’s intervention. Lee and Dey [23, 24] conducted user studies to show that automatically captured photos at a particular frequency by SenseCam [17] (a body-worn camera) are useful in helping older adults with memory impairments recall past events. Other research works have compared SenseCam [17] with written diaries for helping patients with mild cognitive impairment [8], patients with amnesia [27], and older adults with memory impairment [38, 41] and revealed that auto-captured images are more effective than diaries. However, the problem with automatic capture is that if the system does not capture frequently enough, then the specific event or interaction that users want to see may not be in the captured record. These approaches also require users to go through all the contents manually and pinpoint the specific event of interest.

To help users to find content more efficiently, prior research has explored object-based capture systems, which categorize important objects based on hand recognitions or persons based on face recognition in the captured records [3, 25, 26, 28, 42]. For example, Higuchi et al. [16] developed the EgoScanning interface to allow users to navigate to the event of interest using visual cues, such as people or hands. Similarly, Lee and Grauman [26] developed a system that predicts important people or objects automatically from egocentric videos. To provide real-time activity recognition, Lasecki et al. [21] developed Legion:AR which used activity labels collected from crowdsourcing and trained an automatic activity recognition system to recognize future occurrences automatically. Although these approaches reduce the amount of review time, they require continuous recording and storing of video contents, which pose potential privacy concerns [1] in addition to demanding a high volume of storage space and computational power. Another approach is using RFID tags on objects to recognize different activities [35] (e.g., making tea, eating breakfast, and using the bathroom); however, prior research of RFID does not support recognition of the state of objects (e.g., the kitchen stove is on or off, the door is locked or not). Furthermore, some previous approaches require multiple overhead cameras to record the user’s interaction with objects and the state of objects at home. Multiple overhead cameras have a higher cost than a single body-worn camera, and the top-down view angle may not allow the cameras to capture the user’s specific interactions with objects. On the other hand, a body-worn camera creates opportunities to record both direct and indirect user interactions with objects and the states of these objects when they appear in the camera’s field of view.

In this research, we are interested in exploring whether short video clips captured from a body-worn camera every time objects of interest are found within its field of view can help users recall their interactions with the objects and the states of the objects. There are many approaches for determining when an object is seen by a camera. One of the earliest examples of a wearable and/or mobile camera-based device to detect when particular objects are in its field of view is NaviCam [36]. In that work, Rekimoto and Nagao [36] demonstrated applications of recognizing real-world objects by processing video images for color codes. Similarly, we place visual markers near objects of interest in this work as well. However, numerous alternative methods for sensing objects within a camera's field of view have been explored in the literature and presumably can be applied in future versions of the system. For example, Aoki and Matsushita [2] developed Balloon Tag, which used blinking of infrared LEDs to act as invisible markers which transmit infrared patterns to a camera. Patel et al. [34] integrated indoor ultrawideband positioning with a magnetic compass, an accelerometer, a camera, and a laser pointer to allow a mobile device to tracks its own location and orientation within an environment to allow users to interact with fixed-positioned objects from their iCam device's viewfinder. Patel and Abowd [33] demonstrated that information about what located is within a camera's field of view during the point of capture can be used as indices into large video archives to identify the set of video clips which has the objects that the user wants to review later. Hayes and Truong [14] proposed allowing the user to perform the selective archive of captured video segments if it contains information that she wants to store for review later. This approach can reduce the need to store a large amount of unnecessary content and privacy concerns as well. Alternatively, selective archiving can be triggered when particular contexts are sensed. For example, Jiang et al. [20] demonstrated that emotion-driven lifelogging can be supported by using EEG sensors to detect memorial events to the user. In this work, we explore if automatic selective archiving of captured content whenever the pre-selected objects appear within the camera's field of view produces video clips that are both useful and usable for determining the user's interaction history with the objects and the state of these items.

3 FORMATIVE STUDY

3.1 Method

To gain an understanding of what objects or routines older adults need help to keep track of the most, we conducted a formative study with 12 older adults (7 females, 5 males) with the age range from 72 to 80 years old ($\mu = 76.8$, $\sigma = 2.59$). In the formative study, we asked participants about the top five objects or routines that they feel are the most important to remember, and any tools they used to help remember these interactions. The formative interview took approximate 30 minutes. Each participant was compensated 10 CAD for their involvement in the formative study.

3.2 Findings

Based on participant responses from the formative study (Table 1), we learned about two categories of user needs—recalling of the current states of objects and the most recent interactions with objects. Current states of objects include whether or not the door is locked, and the kitchen stove is turned off. Forgetting current states of objects may cause safety hazards or money loss. Remembering past interactions with objects includes scenarios such as when the participant last fed the fish or watered the plants. One of the participants commented:

> *"Remembering the last time that I watered the plants and fed the fish is such a pain to me. I used to write the time down on a piece of paper and leave it near the plants or fish tank. This worked poorly*

> *because sometimes I forgot to write it down. As you know, plants have a basic watering period, and they would die if we do not water them properly. Same for fish, feeding them too much would make them fat and feeding them less may starve them and cause them to die. Plants and fish are not able to complain when not watered well or not fed well."*

Table 1. List of objects that participants would like to keep track of.

If the door was locked
Last time medication was taken
Last time the plant was watered
If various objects (*e.g.,* keys, wallet) were brought when they left home
If the water tap was turned off
If the TV was turned off
If the gas valve was turned off
If the kitchen stove was turned off
If the window was closed
If the lights were turned off
If the food was put back in the fridge
If the circuit breaker was turned off before a trip
Last time the fish was fed

We then asked participants about their current ways of keeping track of both current states of objects and their last interactions with objects. Ten participants indicated a high dependency on paper logs as a reminder. However, they mentioned that writing on paper logs to recall the last time they watered the plants is burdensome when they have different plants which need to be watered at different frequencies. Six participants mentioned that they also relied on their spouse for help. However, participants also claimed the importance of being independent in daily activities. One of the participants commented:

> *"It is really annoying that I have to write down the information about all my plants all the time. Some of them I have to water once per day, some once per month, it is too much work for me. Not only that, I have forgotten to turn off the stove and got my pot burned in the past, so I usually asked my wife to remind me to turn off the stove. However, sometimes she gets tired of remembering these things and sometimes she forgets to remind me. I wish I could memorize these things easily by myself."*

4 SYSTEM

In the formative study, we explored the types of objects (Table 1) that older adults would like to keep track of and their current approaches for remembering these objects and interactions. To investigate whether opportunistically captured video clips could help older adults recall the state of objects and past interactions with objects, we introduce FMT ("Fiducial Marker Tracker"), a real-time capture and access application [39], that runs on a mobile device worn around the neck, to help older adults recall essential information from their daily lives. The system uses the ARToolKit to detect a set of fiducial markers, which are placed near objects that the user wishes to track. Whenever the mobile device's camera detects a marker, the system records a short video clip starting from when the marker is first seen and ending three seconds after the marker is last seen. In the following section, we describe the detailed workflow for using FMT (Fig. 3).

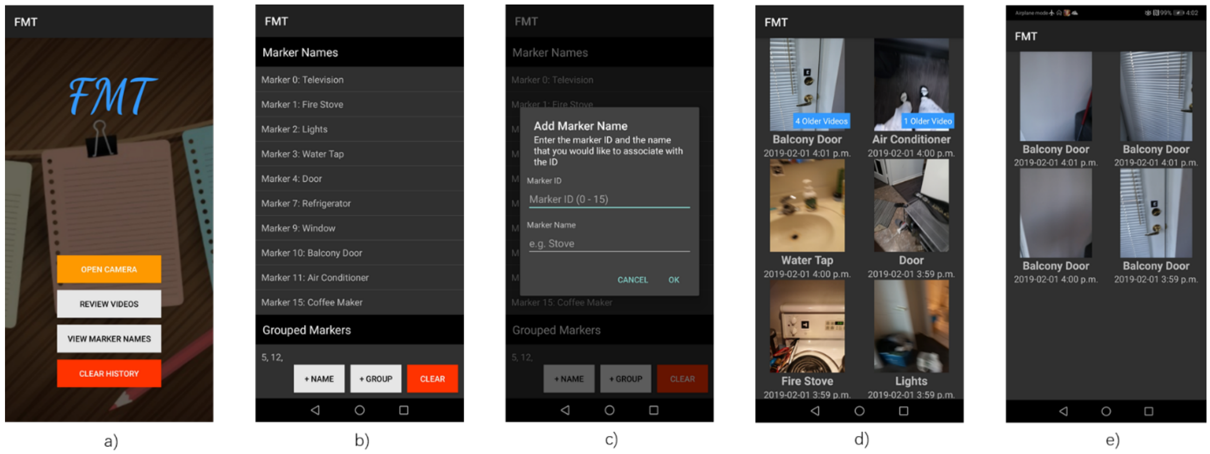

Fig. 3. The FMT application interfaces. (a) Home screen. (b) Interface to add, edit, or delete marker names and marker groups. (c) Popup dialogue to assign a name to a marker. (d) Interface to watch the recorded videos of tracked objects. (e) Older videos of the same objects.

4.1 Setting Up Markers

Before using FMT, the user needs to determine what they would like to track in their homes. Next, the user must stick fiducial markers near what she is interested in tracking, such as a kitchen stove, a door, and plants. Each marker is a 3 x 3 matrix-code marker with a parity-bit added and has a resolution of 72 DPI. The printed markers are 4.8 centimetres in both width and height. Each fiducial marker has a unique ID that can be recognized by the FMT system from four meters when capturing at 640 x 480 pixels.

There are two types of tracking: single marker tracking and grouped markers tracking. Usually, some interactions involve only a single object and require only one marker to be tracked. For example, the user may place a marker just above a door lock to keep track of whether the door is locked. Grouped markers can be set up to track when two or more objects that are typically interacted with together are seen at the same time. The system will only record a video clip when all of the markers in the group are detected in the same camera frame. For instance, when watering a plant (Fig. 4), both the plant and the watering jug must be present in order to complete this activity. Thus, the user can place a marker on the plant base and another on the watering jug, and the video will only be recorded when the camera detects both markers.

Once the markers are set up in the user's home environment, the user can add groups and assign names to the markers on the mobile application. To add a group, the user can enter the marker IDs that she wishes to group together. To assign a name to a marker, the user can enter the marker ID and the name to associate with the marker (Fig. 3(b) and (c)). Assigning names to markers is optional, but it is easier for the user to distinguish between the activities when watching the recorded videos later. By default, if a marker does not have a name, it will be labelled as "Marker X," where "X" is the marker ID. Marker names and grouped markers can be added, edited, and deleted at any time in the View Marker Names interface. Additionally, names can be assigned to markers in camera mode when a marker is detected.

4.2 Storing Video Records

After setting up the markers, the user can launch the application, and wear the mobile device in front of her chest using a strap that hangs from the neck. Then, as the user goes about her daily life, whenever the user passes by a fiducial marker or marker group, the application will capture a video starting from the time that the marker or the marker group is first detected by the camera and ending three seconds after the marker or marker group is no longer within the camera frame. The application records an extra three seconds to capture additional context that might further reveal how the user interacted with the tracked objects (e.g., the room became brighter after turning on the light) (Fig. 4). Moreover, the three seconds buffer allow the system to tolerate accidental occlusions of the marker by the hands or other objects, and the system would still output one single video clip unless the occlusion of the marker is greater than three seconds. If the system recognizes more than one unrelated marker in the field of view, it would use this heuristic to save video clips for each of the markers. The video clip is captured at a resolution of 640 x 480 pixels.

The system will store all past video clips for each marker or marker group. Storing multiple video clips is valuable when the user is interested in several past interactions with the tracked object instead of only the most recent interaction. For example, the user may be interested in knowing how many times she watered the plants in the past few days, rather than just knowing when the last time was that she watered the plants. Furthermore, because the system captures every time it detects a fiducial marker, some of the video records may not contain meaningful interactions between the user and an object. Thus, by storing multiple video records, the system preserves important video clips.

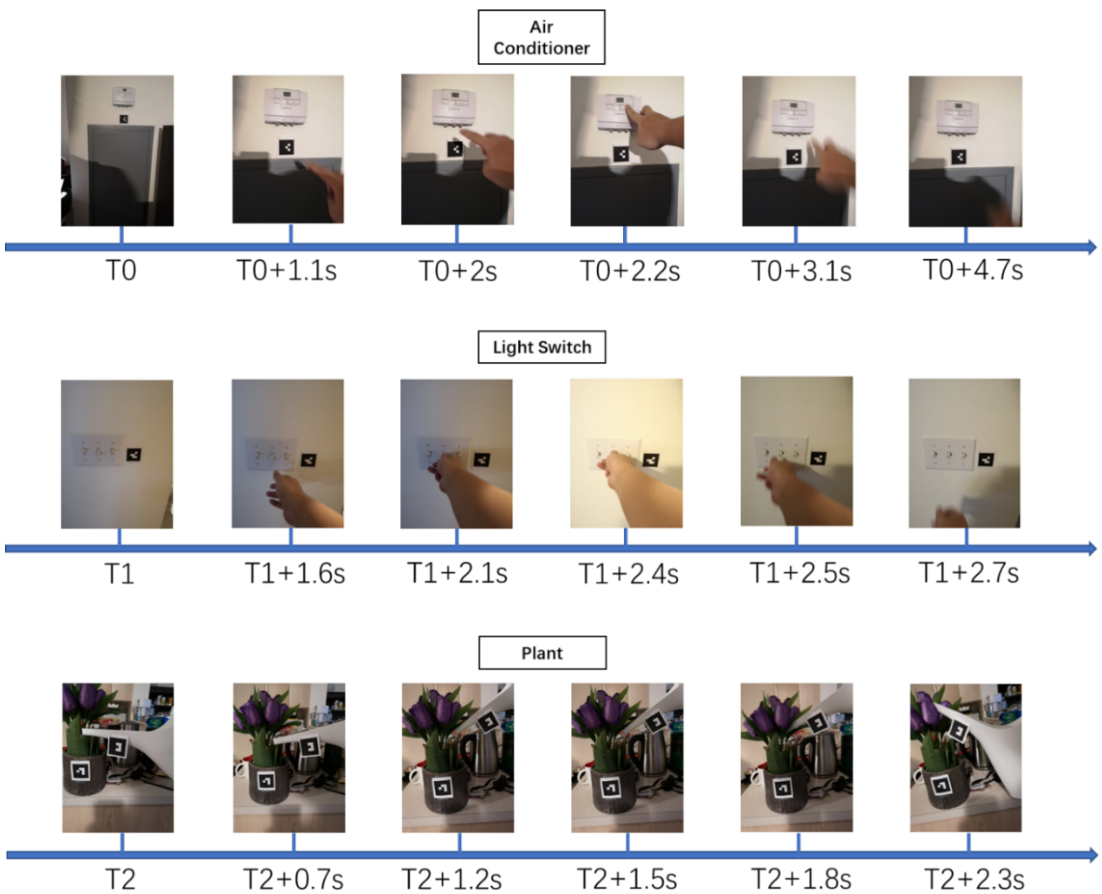

Fig. 4. Example video clips of interactions with different objects captured by FMT: an air conditioner, a light switch, and a plant being watered.

4.3 Reviewing Video Records

The user can watch the recorded videos (Fig. 4) at any time in the application by going to the video reviewing interface (Fig. 3(d)). For each video, the interface shows a thumbnail image (i.e., the middle frame of the video clip) of the recording, the name of the marker that was captured, and the date and time of the recorded interaction. The videos are ordered by timestamp, with the newest appearing at the top. The user can play a video by tapping on the thumbnail. Only the most recent video for each marker is displayed in this main video interface. If a marker has more than one video stored, a label is shown at the bottom right corner of the thumbnail for that marker, indicating the total number of older videos in storage (see Fig. 3(d)). Clicking on the label directs the user to the older videos interface, where the videos can be watched in the same way as those in the primary video interface (Fig. 3(e)).

The main video interface provides the user with quick and convenient access to the last known state of an object. For instance, suppose the user has left her home and is wondering whether or not she has turned off the air conditioner. She can go into the main video interface, tap on the thumbnail of the air conditioner video, and immediately watch her last interaction with the air conditioner (see Fig. 3(d) and Fig. 4). The older videos interface gives a more detailed history of the user's interactions with an object. An example usage scenario is when the user has trouble remembering how many times she has fed the fish in the past three days. She can check the older videos and timestamps of the fish tank to see her recent past interactions with it.

5 EVALUATION

A key finding that we learned from our formative study was that although older adults experience problems, they do not forget the state of different objects at a very predictable and frequent rate. As a result, it would not be practical to conduct a study in which we compare the effectiveness of FMT against other approaches for recalling the state of and interactions with different objects. Instead, we opted to use FMT as a technology probe which allowed us to examine whether video clips opportunistically captured by a body-worn camera when it detects fiducial markers affixed near objects of interest can help older adults keep track of their interactions with these objects and their states as well as why, how, and what older adults would use this approach to recall the state of objects. To do this, we designed and conducted an exploratory user study in which we first set up a controlled environment—a condo apartment—by instrumenting fiducial markers near different objects within a home. Then, we prompted participants to perform a series of activities in which they would directly and indirectly interact with these items while wearing FMT. Afterwards, we asked participants to review video clips captured by FMT to answer questions about whether the application would have been able to help them answer questions about different objects and actions that they have completed. Based on their first-hand experiences with wearing the device and reviewing the captured video clips, we were able to collect their feedback about the approach.

Although a deployment study design would have allowed us to better understand when, why, and how users would actually interact with FMT, it is unclear how long the study would have needed to be in order to capture an adequate number of use cases. Furthermore, because of the potentially unpredictable and infrequent rate at which people forget about the states of and interactions with different objects, it is important that the usability of the system and the efficacy of this approach has been tested so that it can help the users when needed. As a result, we opted to design and conduct the study in the way that is described in the rest of this section rather than as a deployment study.

5.1 Participants

We recruited 12 participants (6 females, 6 males) over 65 years old (μ = 79.17, σ = 3.51) to take part in our study. Participants' education levels ranged from elementary school to undergraduate degrees, and their heights ranged from 145cm to 172cm (μ = 160.08, σ = 8.47) (see details in Table 2). Participants were recruited through posted advertisements on various public and private chat groups of senior communities on Facebook and WeChat. To be eligible for the study, participants were required to have experience with using smartphones. We compensated each participant 20 CAD for their time and involvement in this study.

Table 2. Participant details.

Participant	Gender	Age	Height(cm)
P1	Female	77	162
P2	Male	78	172
P3	Female	75	155
P4	Female	81	150
P5	Male	80	160
P6	Male	83	169
P7	Female	82	145
P8	Male	75	167
P9	Female	84	150
P10	Male	81	166
P11	Male	81	165
P12	Female	73	160

5.2 Scripted Study Design

To conduct this study in a natural, everyday setting, we set up the study in a fifth-floor apartment unit that was intended to feel similar to a home. We scripted a flow that would induce participants to complete a number of activities and interact with different objects during the study. These objects are listed in Table 3, and we chose objects that are representative of the different types of interactions from the formative study results. We prepared an Android mobile device (Huawei P10 Plus with Android Version 7.0) with the FMT application installed. The device was put inside a transparent plastic holder that has a neck strap attached, which was worn around the neck (Fig. 1).

Before the participant arrived, we placed fiducial markers near objects to interact with (e.g., see Fig. 2), and assigned names to these markers in FMT. We prepared the markers and the system beforehand in order to save time. Furthermore, we wanted the set of objects to be consistent for all participants. Hence, we did not ask participants to freely install the markers by themselves before the evaluation.

Two researchers, whom we will refer to as R1 and R2, were involved in the study. R1 acted as the experimenter and R2 acted as a visitor. First, R1 introduced the participant to the apartment and walked the participant through how the various objects in the apartment worked. We asked the participant to consider the apartment as her own home; the participant walked around and familiarized herself with the apartment. Once the participant was ready to proceed, R1 launched the FMT application on a mobile device and asked the participant to wear the device. R1 told the participant that R2 would be coming to visit soon. We instructed the participant to act as they would generally interact with a friend. We did not tell the participant about the purpose of the fiducial markers and the mobile device because we did not want her to pay particular attention to the objects that she will be interacting with later. R1 then left the apartment.

After a short while, R2 knocked on the door. Once the participant let R2 into the apartment, R2 interacted with the participant following a script that would trigger the participant to interact with the objects tagged with fiducial markers. For example, R2 told the participant that the room was too dark, which indirectly prompted the participant to turn on the lights; R2 asked for some tea, which encouraged the participant to boil some water on the stove, and so on. For the final step, R2 told the participant that R1 had asked them to meet downstairs to finish the study. The intention of this last step was to give the participant a reason to leave the apartment and lock the door. The script ended once the participant performed all of the activities. At this point, we stopped running the camera capture on the application. The researchers then invited the participant back into the apartment for the debriefing and post-study interview. During this final part of the study, we asked participants to review videos captured by FMT, answer questions about the usability and usefulness of the captured videos, and provide their feedback about the design and use of FMT.

On average, the study took approximately 20 to 30 minutes. We logged the videos continuously during the user study for the purpose of analyzing frame-by-frame each participant's interactions with various objects from every session.

Table 3. Objects in the apartment that the participant interacted with and the corresponding actions.

Objects	Actions
Water Tap	Turn on/off
Plants (Grouped markers)	Water
Balcony Door	Lock/unlock
Kitchen Stove	Turn on/off
Door Lock	Lock/unlock
Window	Open/close
Lights	Turn on/off
Television	Turn on/off
Air Conditioner	Turn on/off
Coffee Maker	Power on/off
Refrigerator	Store & retrieve food; Open/close

5.3 Post-Study Interview

After the scripted part of the user study, the researchers explained to the participant the purpose of going through the scripted scenario and how FMT worked with the fiducial markers. Next, the researchers conducted a semi-structured interview consisting of three parts to explore the research questions. In part one, we obtained basic demographic information as well as participant's usual practices when leaving their home, such as whether they close the window, turn off the light, and water the plants before leaving home. In part two, we asked questions directly related to the objects that participants interacted with during the study. For each object in Table 3, we asked participants to answer the following 5-point Likert scale questions, and explain their answers:

- Rate whether you feel like you would be able to determine if <the object is in a particular state> (e.g., the door is locked) from the video (from "Very unconfident" to "Very confident")? Why?
- Checking to see if <the object is in a particular state> (e.g., the door is locked) is a problem that you would want to use the system for (from "Strongly disagree" to "Strongly agree")? Why?

We allowed participants to watch videos from the video list freely until they found useful information that would enable them to answer the questions above.

By the third part of the interview, participants had first-hand experiences of wearing FMT and reviewing videos captured by FMT. We were then able to ask the participants 5-point Likert scale questions about their perspectives on various aspects of FMT, such as how they felt about wearing the cellphone in front of the chest (from 1 ("Very uncomfortable") to 5 ("Very comfortable")), and how they felt about wearing the cellphone in front of other people (from "Very embarrassed" to "Very unembarrassed"). Next, we asked them about the privacy concerns that they might have while using FMT. On average, the post-study interview took approximately 30 minutes.

To further understand what older adults would like to keep track of and the way that they would install the markers, we gave each participant three fiducial markers after the exploratory study. We asked participants to place the markers on/close to objects in their own homes that they would use FMT to help them track the most, take pictures of the installed markers, and send those images back to the researchers.

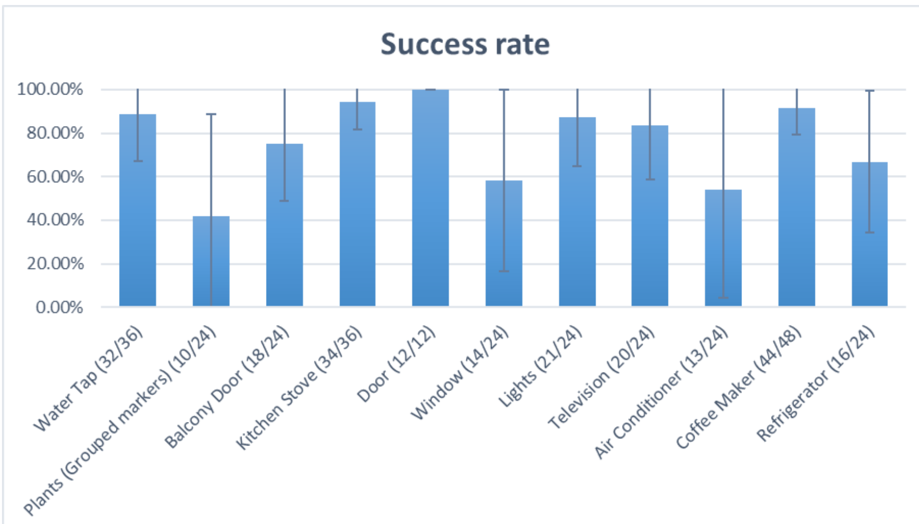

Fig. 5. Success rate and standard deviation of object recognition of FMT.

6 RESULTS

6.1 Do video clips captured opportunistically show the interaction history with & states of objects?

In total, 482 short video clips were captured by the FMT application in the user study, and the average size of a video clip was 1.356 MB (SD = 0.719 MB). In following the user study script, each participant was instructed to directly interact with the listed objects (Table 3) 29 times in total. After analyzing the captured video clips and having the results confirmed independently by two researchers, we found that FMT successfully captured 263 direct interactions out of the 348 scripted interactions in total (29 × 12 participants). Next, we calculated the rate at which FMT successfully captured direct interactions with an object to be 75.6% (263/348) across all participants (SD = 9.9%). The application captured 219 extra video clips in the user study. We found that two reasons caused these extra video clips to be captured. First, hand occlusion and body rotations caused the marker to be out of the camera's field of view for more than three seconds. Thus, there were additional videos captured during a scripted interaction. The other reason is that the application recorded video clips when participants only passed by the marker without interacting with it. Thus, there were additional videos captured opportunistically—which was an effect of the system that we expected to happen and believed would help the system to keep track of the states of objects more accurately (e.g., such as when a light switch was turned off by a partner after the user had turned it on earlier).

The top three objects with the highest fiducial marker recognition success rate were the door (μ = 100%, σ = 0%), the kitchen stove (μ = 94.44%, σ = 12.97%), and the coffee maker (μ = 91.67%, σ = 12.31%) (see Fig. 5). On the contrary, interactions with the plant (Group markers) (μ = 41.67%, σ = 46.87%), the air conditioner (μ = 54.17%, σ = 49.81%), and the window (μ = 58.33%, σ = 41.74%) were recognized with the lowest success rate. We found three major causes for the lower success rate. First, excessively high or low positions of the objects may cause the markers to be out of the camera's field of view. For example, the height of the air conditioner control panel (μ = 54.17%, σ = 49.81%) was about 150 cm, and it was too high to be captured by a camera worn on the chest (see Fig. 6(b)). Second, it was difficult for grouped markers to all appear in the same field of view. In Fig. 6(c), for example, the participant watered the plant and only half of the plant marker appeared in the camera frame; this activity in particular was successfully captured μ = 41.67% (σ = 46.87%) of the time. Third, the white balance of the camera may cause the marker to be less recognizable to the application: the marker colour became darker when the background colour was brighter. For example, the low success rate of the window (μ = 58.33%, σ = 41.74%) was caused by the high luminance of the background and low brightness of the marker (see Fig. 6(a)).

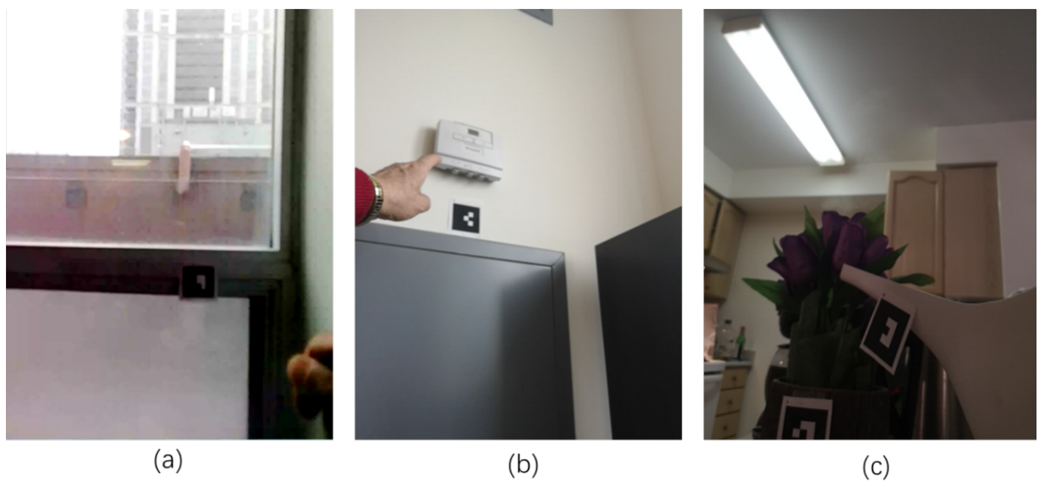

Fig. 6. Some images captured by FMT. (a) Closing the window. (b) Turning off the air conditioner. (c) Watering the plant (Grouped markers).

During the post-study interview, participants were asked whether they feel like they would be able to determine if <the object is in a particular state> (e.g., the door is locked) from the video. Participants answered the question in a 5-point Likert Scale format (from "Very unconfident" to "Very confident") and provided the rationale for their ratings. Participants rated the ability of the video clips to clearly show the state and interaction history of an object as "Confident" (Median = 4, Mode = 5, five is the highest) for all objects. The video clips of the door, the balcony door, and the lights received the highest ratings (Fig. 7).

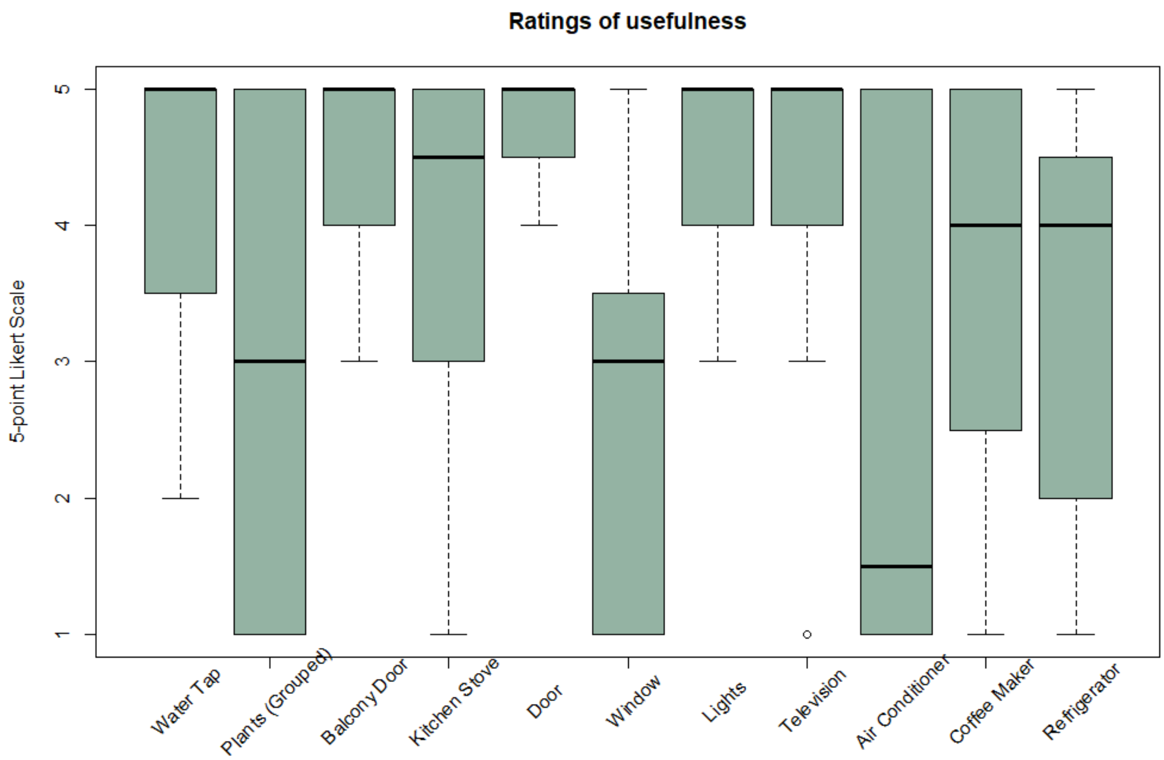

Fig. 7. The ratings for the video clips' usefulness in determining the states of objects. ("1" means "Very unconfident", "5" means "Very confident")

Participants had the lowest ratings for video clips of the air conditioner (Median = 1.5, Mode = 1). One of the reasons was because sometimes interactions with the device were not captured successfully (as mentioned earlier). Additionally, participants reported that the air conditioner panel was too complicated, and the resolution was not high enough to see the buttons and the display clearly (Fig. 4). For example, P3 commented:

> *"I have to spend some time to think about what I did on the air conditioner panel. The panel has too many buttons, and the screen does not have a backlight. It is tough for me to see the letters and the numbers. The resolution of the camera is not clear enough. Oh, by the way, sometimes the orientation is hard for me to see the letters on the panel. Therefore, I think it needs to be improved."*

If we compare the air conditioner with the light switch (Median = 5, Mode = 5), the tracking of the air conditioner required more visual information than the light switch (Fig. 4). Participants needed to see detailed visual contents on the panel to recall past memory. However, participants could determine the state of the light switch simply from the interactions with the actual switch or the changes of brightness (Fig. 4). P2 mentioned the importance of the feedback and background change:

> *"Using background information or feedback helped me to recognize the state of the objects, just like the light, the changing of background brightness could help me understand whether I turned off the light or not. Similarly, I could use smoke or steam to recognize whether I turned off the kitchen stove or not."*

In general, participants felt that they could use the captured video clips to keep track of their interactions and the state for most objects. For example, P5 gave the following positive feedback concerning videos of the door lock (Median = 5, Mode = 5):

> *"I like using this app to help remember whether I locked my door. Sometimes I forget to lock my door, and I have to come back to check. The app is easy to use, and the past video clips are clear when I review interactions with the door lock [in the video]. The past videos also have timestamps on it too. I would not have to worry about the door lock."*

However, participants also suggested some scenarios in which FMT would not be able to help keep track of the objects. For example, multiple people living together can raise some challenges, as pointed out by P9:

> *"I wish the FMT system would allow me to share [the captured videos] across multiple devices. My legs got injured before; it takes me lots of effort to walk around nowadays. Sometimes I turned on the fire and started boiling food. Then I laid down on the sofa and talked to my friends on the phone. Once I think the food is ready, I would ask my husband to turn it off. If I am the only person using FMT, it will not update the state of the stove [on my device because he interacted with it]. Therefore, I want multiple devices to share and have the same video logs."*

This scenario conveys the need for the system to detect and synchronize states of household items from multiple devices worn by different users. With such capability, users can leverage the advantage of living with others to get more frequent and recent updates of the target items. A few participants also mentioned that the height and position of the fiducial markers and the difference in height of a user might cause the FMT system to miss interactions with some objects. For example, P2 mentioned:

> *"I am taller than my wife. There is a height difference of over 20 cm between us. I do not think the fiducial marker at this current height would work properly for both of us. Also, the height of the fiducial marker would affect the performance, just like the air conditioner control panel may be higher than the fridge. I would recommend either placing multiple fiducial markers or having the fiducial markers in different shapes."*

Participants also talked about how objects that are kept inside a shoebox or a drawer would be impossible for this system to keep track of. Therefore, the location of the fiducial markers is a concern that needs to be addressed.

6.2 When would the system be used?

In the interview, many participants mentioned that the reason for potentially using video clips of objects to recall past memories was that they could not pay attention to the state of many different objects at the same time and have a hard time keeping track of things. P5 commented:

> *"I am currently living in an apartment by myself; my son lives far from me. I cannot guarantee that I can keep track of everything by myself. I am easily distracted by other things. Before I go outside for grocery shopping, it always takes me almost 20 minutes to walk around [my home], and I have to make sure everything is turned off. This actually makes my life harder, and I hate to have to keep track of things in my mind."*

To better understand which objects are more crucial for users to use FMT for, we asked participants whether checking to see if <the object is in a particular state> (e.g., the door is locked) is a problem that they would want to use the probe for (from "Strongly disagree" to "Strongly agree"). Participants' ratings and their comments indicated that they would use the system to track objects that may cause extreme consequences (e.g., flooding, fire, theft, and false fire alarm charges). All participants thought that the kitchen stove (Median = 5, Mode = 5) and the door (Median = 5, Mode = 5) were important to use FMT for (see Fig. 8). Some participants had experiences of forgetting to turn off the kitchen stove and causing a fire, or smoke. For example, P4 mentioned in the interview:

> *"I caused a fire several years ago. I remember I was boiling soup in the morning. I realized I need to buy something in the grocery store. Then I walked to the grocery store and forgot to turn off the stove. After I arrived at the grocery store, I total forgot I left the stove on. When I arrived home, the pot was burned really bad, and the smoke was everywhere. Thank god it did not activate the fire alarm of my building; otherwise, I have to pay over 1000 dollars for it."*

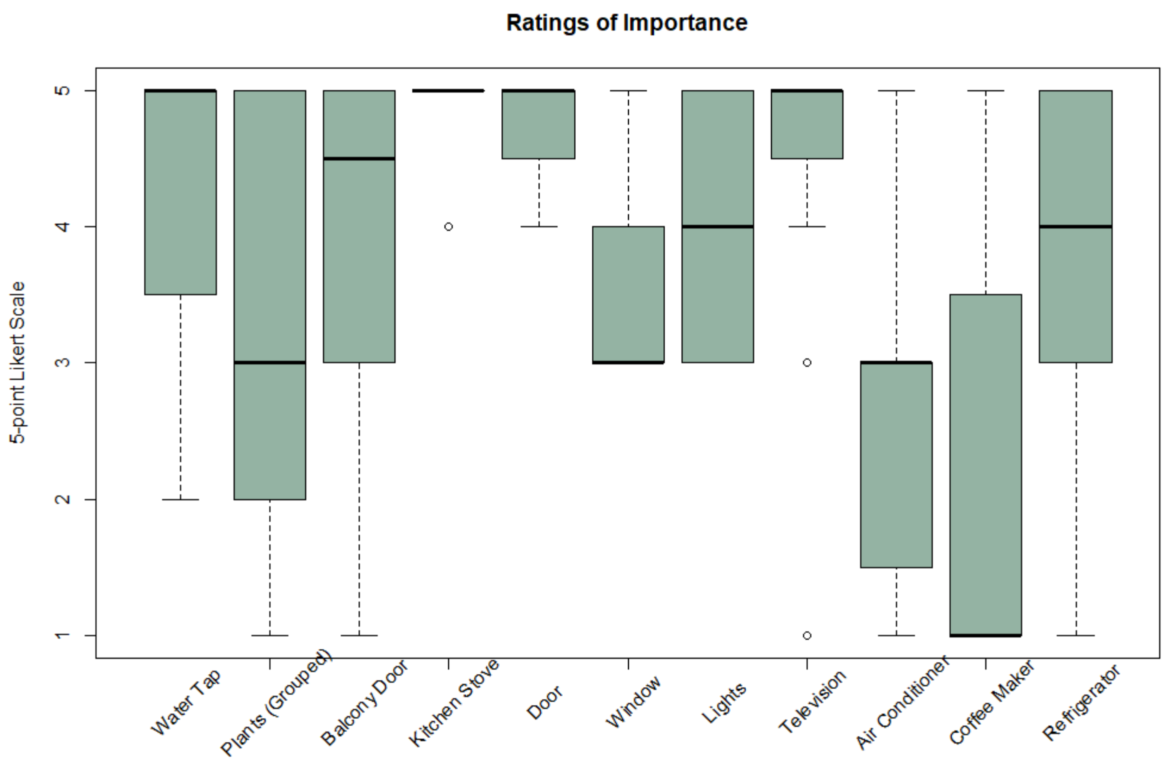

Fig. 8. The rating of importance to keep track of the object. ("1" means "Strongly disagree", "5" means "Strongly agree")

Participants mentioned the importance of saving energy when leaving home. Turning off the light, the television, and the water tap when leaving home would save both water and electricity. In the interview, some participants mentioned that unnecessary power consumption would produce extra cost and negative impact on the environment. P3 commented that:

> *"The electricity and water cost me almost 10% of my income; it would save me some money if I turned off my appliances when I left home. Beyond the money consideration, reducing unnecessary power consumption could reduce the negative effects on the environment. As we know, many people in the world still do not have enough water and electricity in their daily lives. We should save energy for them!"*

We asked participants how they felt about wearing the mobile phone on the chest in the user study. Participants gave answers based on a 5-point Likert scale (i.e., from "Very uncomfortable" to "Very comfortable"). Participants reported feelings between neutral and comfortable (Median = 3.5, Mode = 4) for wearing the cellphone on the chest in daily activities. Participants who felt comfortable mentioned that wearing the cellphone around the neck is similar to wearing a set of keys; however, two participants were worried that wearing the device for longer periods of time may cause fatigue. P6 commented on this concern:

> *"In the scenario study, it is okay for me to wear the cellphone for about 30 minutes; however, I think wearing it for a longer period may make me feel tired. For example, if I wear it for several hours, I may feel tired on my neck. I would recommend accomplishing functionalities on a lighter smartphone, a brooch, or a pair of smart glasses."*

We then asked participants how they would feel about wearing the cellphone in front of other people. Participants answered the question from "Very embarrassed" to "Very unembarrassed." Most participants reported not feeling embarrassed (Median = 5, Mode = 5) to wear the cellphone in public. One of the common reasons was that they wore other stuff in front of their chest already in their daily lives (e.g., a set of keys, ID, and monthly metro pass). P4 commented on this:

> *"I do not feel embarrassed at all. Currently, I wear a set of keys and my ID in front of my chest. The reason for me to do this is that I often forget things when they are in my pocket, and objects are safer when I wear them in front of my chest. I feel that other people are not focused on what I wear in front of my chest when I have my keys and ID."*

We asked participants to discuss what privacy concerns they think they might have while using FMT. Generally, participants were not too concerned about privacy while using FMT. They mentioned two major reasons for it: first, the user is in control of what objects she wants to track, which is different in comparison to automatic lifelogging devices. Second, past video clips are only accessible and reviewed by the user herself. P5 commented:

> *"When I use FMT, I am the person who decides the marker locations. I know the system would not capture unwanted pictures. For example, if I put the marker on the door, it will only capture the video when I lock the door. Even if it captures people's faces when they are walking through, the videos are offline and are only reviewed by myself. I do not think it makes me feel uncomfortable regarding privacy issues."*

6.3 How would the system be used?

The FMT application saved all past video clips of each object, and these video clips were presented in the order of the interaction date and time (from "Newest" to "Oldest"). FMT saved the video clips of objects when users were near them and the fiducial markers were detected within the field of view of the camera. In the user study, we found that participants recorded more short video clips than the times of interactions with objects. In this section, we examine whether the most updated video clips captured opportunistically could be used to keep track of the interaction history and states of objects, and how many video clips and how much of a clip participants need to review for recall. Therefore, we logged the behaviours of participants while they interacted with FMT during the interview.

From the participants' behaviours in the study, we discovered that they checked the latest video clips the most (μ = 78.79%, σ = 11.19%). For objects like the kitchen stove and TV, all participants checked the most updated clip and did not continue to check the old clips. Participants revealed that since the most recent video clip already showed the state of the kitchen stove or the TV, they did not need to see the actual interaction. P5 commented on this:

"Well, I do not need to know when I turned off the stove and my previous interactions with the stove. If I can make sure there is no fire on the stove, that is all I need, and that is all I care about. That is the reason why I just reviewed the last video clip of the stove."

However, for some objects, participants checked multiple video clips until they found what they wanted to see. In particular, for the air conditioner (Fig. 4), one of the participants found what she wanted by the fourth most recent video clip. P9 commented on this:

"Checking the video clips of objects to determine the states of objects are different based on what the object is. I found for the stove I just need to find whether the fire was off the last time I saw it. However, I find that the video clips of simply passing by the air conditioner and door lock without interaction information are useless to me."

In the 30-minute interview, we found that participants spent an average of 149.25 seconds (σ = 12.23 seconds) using the user interface and reviewing the video clips. When participants reviewed a video clip, on average they watched each for 9.17 seconds (σ = 4.29 seconds). We found that different objects required different numbers of video clips to show all of the information that participants needed. If the state of an object could be easily determined even without seeing the actual interaction with it, participants did not need to review many video clips; for example, participants spent an average time of 8.91 seconds (σ = 2.02 seconds) using the system and reviewing the video clips of the kitchen stove. However, participants usually reviewed more video clips for objects when they needed to see the actual interactions (e.g., the light switch (μ = 21.42 seconds, σ = 3.55 seconds)).

Fig. 9. Pictures of fiducial markers with objects captured by participants.

6.4 What objects would older adults use opportunistically-captured video clips to keep track of?

After participants gained experience with FMT through the user study, we wanted to better understand how participants would install the fiducial markers and what objects they would like to track in their most familiar environment (e.g., their own home). We gave each participant three fiducial markers and asked the 12 participants to place the markers on their appliances or objects they would use FMT to help them track the most, take pictures of the installed markers, and send those images back to the researchers (see Fig. 9). In Table 4, we listed different objects captured by participants and the number of images of each type of object we received. Similar to the setup in our controlled study, participants preferred to use the fiducial markers to track the door locks (10), and the kitchen stoves (7) (see Table 4). In addition, we discovered that participants also placed the fiducial markers close to the power plugs of appliances (Fig. 9(i) and (l)). We asked one of the participants why she placed the fiducial marker near the power plugs. The participant replied that she did not trust the "sleep" or "off" mode of her appliances, and she always unplugged the power from the wall when she did not use it. These appliances include ovens, pressure cookers, rice cookers, and toasters. She commented:

> *"I do not understand how the inside part (circuit) works in my appliances. That is why I feel it is safer to just unplug the power from the wall. I know this will not cause a fire or break the device for sure. I found that FMT could help me track all my plugged-in power cables and reduce safety concerns."*

In terms of the users' behaviours on marker placement, we learned that participants mostly placed markers either to the left (13) or to the right (11) of the objects that they would like to keep track of (Table 4). This is most likely due to the limited access to the top or bottom of objects (e.g., the door, and the kitchen stove). In addition, participants placed the fiducial markers directly on the refrigerator (Fig. 9(n)), the laundry machine (Fig. 9(k)) and the television to track states of objects and past interactions. In terms of the alignments, three photos that we received showed that the fiducial markers were not axis aligned relative to the objects (e.g., Fig. 9(l)). Finally, we found that all participants placed the fiducial markers where there was no visual obstacle, and they preferred to place the markers close to the tracked objects.

Table 4. Summary of the photos taken by 12 participants and positions of the fiducial markers relative to the objects to track.

Object	Num of pictures	Right	Left	Above	Below	On
Door	10	4	5	1	0	0
Fire Stove	7	3	3	1	0	0
Power Plug	5	2	2	1	0	0
Air Conditioner	3	0	1	1	1	0
Light	3	1	1	1	0	0
Refrigerator	2	0	0	0	0	2
Television	1	0	0	0	0	1
Window	1	0	0	0	0	1
Laundry Machine	1	0	0	0	0	1
Medication	1	1	0	0	0	0
Water Tap	1	0	0	1	0	0
Balcony Door	1	0	0	1	0	0

6.5 Power Consumption

We also recorded the battery consumption (Table 5) of the smartphone throughout the duration of the study. We ensured that FMT was the only application opened (and WIFI and Bluetooth were off) while running the study. We found that the cellphone consumed about 3.17% of the battery for an average of 23 minutes per participant during the user study. Because we logged the video continuously during the user study for the purpose of analyzing frame-by-frame each participant's interaction with various objects for every session, the power consumption is likely to be lower in a real usage scenario where the system is running without the continuous video logging.

Table 5. Battery consumption of the smartphone during the user study.

Participant	Study Duration	Start Battery	End Battery	Consumption
P1	20 Minutes	49.00%	44.00%	5.00%
P2	24 Minutes	84.00%	80.00%	4.00%
P3	26 Minutes	30.00%	25.00%	5.00%
P4	23 Minutes	97.00%	94.00%	3.00%
P5	22 Minutes	91.00%	88.00%	3.00%
P6	21 Minutes	84.00%	82.00%	2.00%
P7	23 Minutes	76.00%	73.00%	3.00%
P8	29 Minutes	100.00%	99.00%	1.00%
P9	23 Minutes	85.00%	82.00%	3.00%
P10	20 Minutes	90.00%	87.00%	3.00%
P11	23 Minutes	84.00%	81.00%	3.00%
P12	22 Minutes	95.00%	92.00%	3.00%
Average	23 Minutes	80.42%	77.25%	3.17%

7 DISCUSSION AND LIMITATION

7.1 Personalization & Installation

Personalizing the system to work for different individuals is an important but difficult challenge. Our participants were of different heights and different mobility conditions. A setup that captured fiducial markers well for one user did not always capture markers clearly for others due to their height differences. Although it is possible to consider other ways to position the markers, it is really the objects that need to be detected and recorded in the video clips. When the objects are located in spots that are hard to be captured by a body-worn camera, then it is challenging to record useful clips. Thus, an important consideration that must be explored further is the camera design itself and how to place it on the user's body—regardless of their height and mobility—so that it can capture objects that may not be easily captured by a device worn around the neck.

One of the limitations of the reported study is that we did not examine the users' ability to install and setup FMT by themselves (e.g., setting up marker names and grouped markers). It is necessary to explore the setup barriers involved in using FMT in the future. The installation of the system must also be considered because it will ultimately be used in a real-life setting. For the purposes of our study, we installed the fiducial markers ahead of time and manually ensured that the markers were placed in a way that increased the chance that video clips of the objects would be automatically recorded as the participants walked around. After the interviews, we asked participants to install fiducial markers on objects in their own homes. We analyzed their preferences in marker installation, and the relative positions between the objects and the fiducial markers. However, it is unclear whether markers installed by older adults could keep track of their desired objects. Afterwards, they must verify that the system would be able to successfully track particular objects, know when and why the system is not able to track certain objects well, and make adjustments. Addressing the installation challenge requires an understanding of how people would install the markers, wear the capture device, and their expectations at each step.

7.2 Wearable Form Factors

Participants were asked to wear a smartphone for 20-30 minutes in our study. The smartphone simply acted as a representative body-worn camera which would selectively capture first-person video clips of past interactions that the user has with objects. In that regard, participants imagined it would be beneficial to wear such a device in their own lives. However, participants mentioned the possibility of getting fatigue from wearing a smartphone around their neck for the whole day. This was also another limitation of the study, because the study design did not examine whether users would actually wear a device with this form factor throughout the day on a regular basis. On the other hand, the current configuration assumes that the user would already have a smartphone that can be leveraged for capture in this way. However, the user would have to dedicate their smartphone to capturing video clips. This means that when the user interacts with an object that she wants the system to help track, it would be difficult for her to use the smartphone for any other purpose. If the user interacts with the phone in any other way—this includes reviewing past videos captured by FMT—then the system may not detect fiducial markers properly nor record new video clips accordingly during this time. Thus, it is important to explore alternative body-worn camera solutions. Ideally, such a solution would not burden the user with the need to carry an additional device with them throughout the day. For example, P2 and P8 mentioned that they often wear glasses at home. For such users, their glasses can be augmented with a camera to support the opportunistic capturing of their interactions with objects in their home.

7.3 Interface Design, Storage Considerations & Battery Concerns

In our study, the system tracked 11 objects tagged by different fiducial markers. Users may end up wanting to track many more objects in practice, especially when there are multiple locations involved. Being able to quickly find and review the interaction history and the state of a specific object amongst many objects will be an essential challenge for users.

Participants mentioned that they placed varied amounts of importance in different objects. They considered turning off the kitchen stove and locking the door as the most crucial memories to recall. FMT currently uses a time-based interface to show the captured data – the most recent interaction appears on the top (see Fig. 3(d)). However, the interface could also be designed as a rank-based system in which users determine what objects they would always like to see at the top. P4 claimed in the interview:

> *"I do not care about remembering to turn off the light every day. I want the system to always show crucial objects on top. Compared to [potentially] causing a fire or flooding, wasting some energy means nothing to me."*

FMT may also be applied for beyond-home use scenarios—such as offices, cars. Therefore, how to design the interface to track multiple locations is essential to users. Should the system have multiple interfaces for various locations? How should one determine the coverage of a location (e.g., a room or an entire condo)? Perhaps it is vital to consider whether the system should have a rank of importance for various locations (e.g., should the interactions at home be considered more important than at the office?).

The study software was also designed to keep all of the video clips captured for each participant until it was explicitly removed. Overall, the FMT system captured a total of 482 video clips, which took up 653.6 MB in storage space, during the study across all of the participants. Of the 482 video clips, we found 219 clips were unnecessary (e.g., did not offer any new information over a previously captured clip) or were unusable (e.g., the object itself was occluded in the video although the marker was visible, or it did not show the object clearly within the captured video). Although each clip was on average only 1.356 MB, there were approximately 18 such video clips captured by the system for each participant during the 20-30 minutes period when they wore the device. When considering instances in which the user would potentially wear FMT continuously throughout the day and the number of objects being tracked increases to more than the 11 used in the study, the amount of unnecessary and unusable video clips—and storage space wasted by the system to store them—could grow to be unwieldy over time. Thus, an important issue that must be explored next is how to design the system so that it keeps only the most useful video clips while deleting unnecessary ones so that the system does not consume a lot of storage space.

In the current implementation, we did not explore ways to automatically remove unnecessary video clips. Some simple approaches to managing the size of the storage space used by FMT is to either delete video clips older than a particular number of days or to place a limit on the number of video clips that would be kept for each object. However, a time-based approach might mean that important video clips showing the last known state for infrequently used objects will be discarded when they should be kept. A count-based approach might mean that FMT would store many video clips captured when users only pass by objects in high-traffic areas; those videos might show the most recent state of these objects but the user may not be able to review the specific interaction history related to these items if those clips have been removed automatically. Thus, the system must be designed to learn how to appropriately store and discard video clips for different objects.

Finally, we found that approximately 3.17% of the battery was consumed for an average of 23 minutes of usage in the user study. If a user wears the device for 8 hours, this app alone would consume about 66.2% of the battery. However, other apps and resources on the phone, such as the WIFI module, the Bluetooth module, the phone screen, etc. will also consume the battery. Thus, it is possible that the battery would be drained in under 8 hours. Future work should examine the effects of different capture parameters on power consumption to identify the optimal settings to use with the FMT application so that it will not drain the battery so much. Moreover, developing an algorithm that automatically activates or deactivates capture using context (e.g., activate when the user has entered the home, deactivates it after they have left home) can help to reduce the consumption of the battery as well.

7.4 Multi-user Scenarios

In our study, FMT was used by only one participant at a time. However, as pointed out by one participant (P11), the design of the FMT system must take into consideration situations in which users live with others:

> *"Currently, FMT was only used by myself. I do think all my family members need it. When someone interacted with an object, such as the kitchen stove, or the TV, every family member should all receive the update. For example, let's assume that I am the only person who used FMT at home. If I turned on the stove and someone else turned it off for me, how can I know that from the system? So I think it is definitely useful and important to support multiple devices to share the same video clips in the same system."*

This points out the need for a system to detect and synchronize states of household items from multiple devices worn by different users. In this way, users would be able to leverage the advantage of living with others to get more frequent updates of the tracked items. On the other hand, some privacy concerns may arise because of this sharing. For example, captured video clips show when a user has interacted with different objects or has completed an action. Through this, the video clips are indirectly showing contextual information about a person (e.g., the user's presence at home, what they might be doing, and so on). An important design consideration is how to prevent the system from unintentionally revealing information about different users that it should not.

7.5 Proactive Reminder & Smarter System

FMT was designed as a system to help participants recall past interactions with objects when they forget the states of objects. The user currently has to review the video clips to check the states of objects. However, *"why does the system require checking?"* was a question that participants brought up during the user study. One of the limitations of FMT is that it does not provide proactive reminders to the user when she interacts with various objects. From our study, we believe that proactive reminders would be helpful in several scenarios. For example, a proactive reminder if the user forgot to lock the door when leaving home would be more helpful than requiring the user to check the video clips later at the workplace or somewhere else. Furthermore, proactive reminders for different objects can help to reduce the mental burden associated with checking the video clips on the system manually. Such a system must be designed to learn and detect the typical states of objects. For example, computer vision can be used to analyze and learn how different objects look before and after any direct interactions that the user has with them, as well as how long objects are left in particular states. When an object is left in a state for longer than normal, the system can proactively remind the user about these deviations. FMT also allowed the user to track interactions with multiple objects simultaneously through grouped markers. Adding a way to detect interactions with a group of markers in a sequential manner could be a potential way to implement proactive reminders. For example, locking the front door can be decomposed into several steps. The user has to first open the door from the inside, step out of the home, then close and lock the door from the outside. We could place one fiducial marker on the inside of the door, and another one on the outer door lock. The door-locking activity is only considered complete if the system detects the inner marker, followed by the outer marker. If the system only detected the inner marker, then it should recognize that the user is leaving home, and proactively remind the user to lock the door if it fails to detect the outer marker after a certain time threshold.

Another way of supporting proactive reminders is to build a system that learns the user's routines and tracks the user's interactions with objects. For instance, the system can learn that the user typically feeds the fish twice per day. If the user has fed the fish twice already, the system would proactively remind the user of this if she tries to feed again that day. Similarly, the system could remind the user about the need to feed the fish on time. In general, a smarter system with proactive reminders may help to reduce the mental burden of checking the state of objects manually.

8 CONCLUSION

In this paper, we evaluated whether video recordings of objects opportunistically captured from a neck-worn device could help older adults keep track of the states of these objects and their interaction history. We conducted a formative study with 12 older adults to understand objects or routines that they forget the most. We then conducted a study with another 12 older adults aged over 65 using FMT as a technology probe to investigate whether opportunistically captured video clips would be able to show them states of objects or their interaction history with objects, and what, why, and how they would use such a system in their lives. Our results showed that video clips captured from a body-worn camera recorded 75.6% of the direct interactions that participants had with the tagged objects. Although there were several different reasons which prevented the system from successfully capturing all of the interactions with the objects in the study, the system was able to capture video clips each time users were near the objects and the fiducial markers were within the field of view of the camera. These opportunistically captured clips helped to compensate for some instances in which the application was not able to record direct interactions. Overall, the study showed FMT is a potentially useful and usable memory aid, and that additional features such as different ways to sort and show the recorded videos and proactive reminders could be added to future versions of the software.

9 REFERENCE

[1] Allen, A.L. 2008. Dredging up the past: Lifelogging, Memory, and Surveillance. *The University of Chicago Law Review*. 75, 1 (2008), 47–74. DOI:https://doi.org/10.2307/20141900.

[2] Aoki, H. and Matsushita, S. 2002. Balloon Tag: (in)visible marker which tells who's who. *ieeexplore.ieee.org* (2002), 181–182.

[3] Arev, I., Park, H.S., Sheikh, Y., Hodgins, J. and Shamir, A. 2014. Automatic editing of footage from multiple social cameras. *ACM Transactions on Graphics*. 33, 4 (Jul. 2014), 1–11. DOI:https://doi.org/10.1145/2601097.2601198.

[4] Block, L.G. and Morwitz, V.G. 1999. Shopping lists as an external memory aid for grocery shopping: Influences on list writing and list fulfillment. *Journal of Consumer Psychology*. 8, 4 (Jan. 1999), 343–375. DOI:https://doi.org/10.1207/s15327663jcp0804_01.

[5] Boldt, R. 2013. The Last-Seen Image : an Image-Based Approach for Finding Lost Objects Using a Head-Mounted Display. 91 (2013).

[6] Brewer, W.F. 1988. Qualitative analysis of the recalls of randomly sampled autobiographical events. *Practical aspects of memory: Current research and issues, Vol. 1: Memory in everyday life*. 263–268.

[7] Brewer, W.F. 1986. What is autobiographical memory? *Autobiographical memory*. c (1986), 25–49. DOI:https://doi.org/10.1017/CBO9780511558313.006.

[8] Browne, G., Berry, E., Kapur, N., Hodges, S., Smyth, G., Watson, P. and Wood, K. 2011. SenseCam improves memory for recent events and quality of life in a patient with memory retrieval difficulties. *Memory*. 19, 7 (2011), 713–722. DOI:https://doi.org/10.1080/09658211.2011.614622.

[9] Cavanaugh, J.C., Grady, J.G. and Perlmutter, M. 1983. Forgetting and Use of Memory Aids in 20 to 70 Year Olds Everyday Life. *The International Journal of Aging and Human Development*. 17, 2 (Sep. 1983), 113–122. DOI:https://doi.org/10.2190/H7L2-K3XK-H32K-VW89.

[10] Centers for Disease Control and Prevention 2012. Identifying Vulnerable Older Adults and Legal Options for Increasing Their Protection During All-Hazards Emergencies: A Cross-Sector Guide for States and Communities. *U.S. Department of Health and Human Services*. (2012).

[11] Dalgleish, T. and Cox, S.G. 2002. Memory and emotional disorder. *The handbook of memory disorders*. (2002), 437–457.

[12] Ellis, D.P.W. and Lee, K. 2004. Minimal-impact audio-based personal archives. *Proceedings of the the 1st ACM workshop on Continuous archival and retrieval of personal experiences - CARPE'04* (2004), 39.

[13] Hayes, G.R. 2004. The Personal Audio Loop: Designing a Ubiquitous AudioBased Memory Aid,? BT - Proc. 6th Int?l Conf. Human Computer Interaction with Mobile Devices and Services. (2004), 168–179.

[14] Hayes, G.R. and Truong, K.N. 2009. Selective Archiving: A Model for Privacy Sensitive Capture and Access Technologies. *Protecting Privacy in Video Surveillance*. Springer London. 165–184.

[15] Hertzog, C. & Dixon, R.A. 1994. Metacognitive development in adulthood and old age. *Metacognition: Knowing about Knowing*. J. Metcalfe & A. Shimamura, ed. 227–251.

[16] Higuchi, K., Yonetani, R., Sato, Y. and Higuch, K. 2017. EgoScanning: Quickly Scanning First-Person Videos with Egocentric Elastic Timelines. (2017). DOI:https://doi.org/10.1145/3025453.3025821.

[17] Hodges, S., Williams, L., Berry, E., Izadi, S., Srinivasan, J., Butler, A., Smyth, G., Kapur, N. and Wood, K. 2006. SenseCam: A Retrospective Memory Aid. (2006), 177–193. DOI:https://doi.org/10.1007/11853565_11.

[18] Information, C.I. for H. 2007. Health Care in Canada, 2011. *Reactions Weekly*. NA;, 1157 (2007), 3. DOI:https://doi.org/10.2165/00128415-200711570-00010.

[19] Jebara, T., Schiele, B., Oliver, N. and Pentland, A. 1998. DyPERS: Dynamic personal enhanced reality system. *In Proc. 1998 Image Understanding Workshop*. (1998).

[20] Jiang, S., Zhou, P., Li, M.O., Li, Z. and Li, M. 2019. Memento: An Emotion-driven Lifelogging System with Wearables. *ACM Trans. Sen. Netw*. 15, 8 (2019). DOI:https://doi.org/10.1145/3281630.

[21] Lasecki, W.S., Song, Y.C., Kautz, H. and Bigham, J.P. 2013. Real-time crowd labeling for deployable activity recognition. *Proceedings of the 2013 conference on Computer supported cooperative work - CSCW '13* (2013), 1203.

[22] Le, H., Clinch, S., Sas, C., Dingler, T., … N.H.-P. of the and 2016, undefined Impact of video summary viewing on episodic memory recall: Design guidelines for video summarizations. *dl.acm.org*.

[23] Lee, M.L. and Dey, A.K. 2008. Lifelogging memory appliance for people with episodic memory impairment. *Proceedings of the 10th international conference on Ubiquitous computing - UbiComp '08* (2008), 44.

[24] Lee, M.L. and Dey, A.K. 2007. Providing good memory cues for people with episodic memory impairment. *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility - Assets '07* (2007), 131.

[25] Lee, Y., Ghosh, J., Pattern, K.G.-C.V. and and 2012, U. Discovering Important People and Objects for Egocentric Video Summarization. *ieeexplore.ieee.org*.

[26] Lee, Y.J. and Grauman, K. 2015. Predicting Important Objects for Egocentric Video Summarization. *International Journal of Computer Vision*. 114, 1 (Aug. 2015), 38–55. DOI:https://doi.org/10.1007/s11263-014-0794-5.

[27] Loveday, C. and Conway, M.A. 2011. Using SenseCam with an amnesic patient: Accessing inaccessible everyday memories. *Memory*. 19, 7 (2011), 697–704. DOI:https://doi.org/10.1080/09658211.2011.610803.

[28] Lu, Z. and Grauman, K. 2013. Story-driven summarization for egocentric video. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Jun. 2013), 2714–2721.

[29] Maylor, E.A. 1993. Aging and forgetting in prospective and retrospective memory tasks. [erratum appears in Psychol Aging

1995 Jun;10(2):190.]. *Psychology & Aging*. 8, 3 (1993), 420–428.

[30] Mynatt, E.D. and Rogers, W.A. 2001. Developing technology to support the functional independence of older adults. *Ageing International*. 27, 1 (2001), 24–41. DOI:https://doi.org/10.1007/s12126-001-1014-5.

[31] Niforatos, E., Cinel, C., Cathleen, M.C., Langheinrich, M. and Ward, G. 2017. Can Less be More? Contrasting Limited, Unlimited, and Automatic Picture Capture for Augmenting Memory Recall. *IMWUT*. 1, 2 (2017). DOI:https://doi.org/10.1080/07366981.2012.645389.

[32] Nolan, B.A.D., Mathews, R.M. and Harrison, M. 2001. Using external memory aids to increase room finding by older adults with dementia. *American Journal of Alzheimer's Disease and other Dementias*. 16, 4 (Jul. 2001), 251–254. DOI:https://doi.org/10.1177/153331750101600413.

[33] Patel, S.N. and Abowd, G.D. 2010. The ContextCam: Automated Point of Capture Video Annotation. *Springer*. 301–318.

[34] Patel, S.N., Rekimoto, J. and Abowd, G.D. 2006. iCam: Precise at-a-Distance Interaction in the Physical Environment. *Springer*. 272–287.

[35] Patterson, D.J., Fox, D., Kautz, H. and Philipose, M. *Fine-Grained Activity Recognition by Aggregating Abstract Object Usage*.

[36] Rekimoto, J. and Nagao, K. *The World through the Computer: Computer Augmented Interaction with Real World Environments*.

[37] Sugawara, E. and Nikaido, H. 2014. *World Report on Ageing and Health*.

[38] Tan, D., Berry, E., Czerwinski, M., Bell, G., Gemmell, J., Hodges, S., Kapur, N., Meyers, B., Oliver, N., Robertson, G. and Wood, K. 2007. Save everything: supporting human memory with a personal digital lifetime store. *Personal information management*. 90–107.

[39] Truong, K.N., Abowd, G.D. and Brotherton, J.A. 2001. Who, What, When, Where, How: Design Issues of Capture & Access Applications. 209–224.

[40] Verbrugge, L.M. 1980. Health diaries. *Medical Care*. 18, 1 (1980), 73–95. DOI:https://doi.org/10.1097/00005650-198001000-00006.

[41] Woodberry, E., Browne, G., Hodges, S., Watson, P., Kapur, N. and Woodberry, K. 2015. The use of a wearable camera improves autobiographical memory in patients with Alzheimer's disease. *Memory*. 23, 3 (2015), 340–349. DOI:https://doi.org/10.1080/09658211.2014.886703.

[42] Xu, J., Mukherjee, L., Li, Y., Warner, J., Rehg, J.M. and Singh, V. Gaze-enabled Egocentric Video Summarization via Constrained Submodular Maximization. *cv-foundation.org*.

[43] Yasami, M., Dua, T., Harper, M. and S, S. 2013. Mental Health of Older Adults, Addressing a Growing Concern. *Mental health and older people* (2013), 4–9.