Institute of Visualization and Interactive Systems

University of Stuttgart
Universitätsstraße 38
D–70569 Stuttgart

Master's Thesis Nr. 29

Development and Evaluation of Automatic Video Recaps from Lifelog Data

Huy Viet Le

Course of Study:	Softwaretechnik
Examiner:	Jun.-Prof. Dr. Niels Henze
Supervisor:	Dipl.-Medieninf. Tilman Dingler, Dr. Corina Sas (external)
Commenced:	13 April 2015
Completed:	13 October 2015
CR-Classification:	H.5.m

Kurzfassung

Lifelogging Kameras sind kleine, tragbare Kameras, die bis zu 1500 Bilder am Tag aufnehmen. Verwandte Arbeiten haben gezeigt, dass diese nicht nur gegen episodische Gedächtnisstörungen helfen, sondern auch als effektive Gedächtnisstützen für die allgemeine Bevölkerung eingesetzt werden können. Die große Menge an Bildern erschwert jedoch einen Einsatz als effektive Gedächtnisstütze ohne dabei einen erheblichen Aufwand in Kauf zu nehmen. Es wäre also wünschenswert, dass relevante Bilder automatisch herausgefiltert und in einer angemessenen Form präsentiert werden, sodass der Abruf von episodischen Erinnerung gefördert wird. Aus diesem Grund entwickeln wir in vorliegender Arbeit eine Software, die relevante Bilder erkennt und diese in Form einer Videozusammenfassung präsentiert. Wir führen dazu eine fünfwöchige Studie durch um Anforderungen für solche Videozusammenfassungen zu sammeln und zu evaluieren. Kriterien für solche relevanten Bilder werden in dieser Arbeit vorgestellt, wobei auch auf die Präsentationstechnik der Bilder eingegangen wird, die das episodische Gedächtnis bestmöglich unterstützen sollen. Wir evaluieren zudem unsere Videozusammenfassungen und konnten dabei zeigen, dass es im Vergleich zu nicht-filternden Ansätzen (u.a. Zeitraffer bzw. die manuelle Betrachtung aller Bilder) keinen signifikanten Unterschied in der Effektivität als Gedächtnisstütze gibt. Darüber hinaus bevorzugten Probanden unsere Videozusammenfassungen gegenüber den nicht-filternden Ansätzen aufgrund einer besseren Nutzerfreundlichkeit, was vor allem für die Beförderung dieser Techniken vom klinischen Bereich in den Alltag eine große Rolle spielen könnte.

Abstract

Lifelogging cameras are small and wearable cameras that capture up to 1,500 images per day. Prior work has shown that these images support the episodic recall of not only the memory-impaired patients but also of the general population. However, the sheer volume of the captured image sets exceeds the capability of users to review them on a regular basis. Hence, it would be desirable to automatically detect relevant images in a set of captured images and present them in a way that supports the episodic recall. For that reason, we develop a software that recognizes relevant images and present them in the form of a video summary. Requirements for these video summaries were elicited and evaluated in the context of a five-week study. In this work, we present criteria for relevant images and how they should be presented to benefit the episodic recall. An evaluation of our video summaries revealed that there is no significant difference in the effect on the episodic memory in comparison to review methods that present the entire lifelogging image set. Moreover, participants prefer video summaries over said non-summarizing review methods due to a better usability which can play an important role in elevating this memory augmentation technology from a clinical niche application to a mainstream technology.

Contents

List of Figures

List of Tables

List of Algorithms

1. Introduction

Recordings such as diaries, notes, protocols, photographs, and other analog records have always been useful cues to preserve and reflect on memories. In the past, these were recorded analogous which required a huge effort. With today's technology, most of them can be recorded automatically. Lifelogging systems such as fitness trackers, wearable cameras and mobile applications that track visited places, users behavior or help to create daily journals are getting more known and powerful due to the rise of wearable and mobile computing. These inventions render the creation of lifelogging records to a viable daily practice. It stands to reason to use this tremendous data set to augment the rather oblivious human memory. Previous research work has already investigated wearable cameras (i.e. lifelogging cameras) such as the Microsoft SenseCam[1] or the NarrativeClip[2] as memory aids and presented a successful use in supporting memory impaired patients [LD07, KHBS10, BBK$^+$11, LD08, YKK$^+$09, WBH$^+$15].

The RECALL project[3] is a European research project that aims to "*elevate memory augmentation technologies from a clinical niche application to a mainstream technology*". This thesis has been partly conducted in the context of RECALL with the vision of using images captured with lifelogging cameras as an effective and effortless memory augmentation to support users with no memory impairments to recall past events, also called *reminiscence*.

Prior research has already investigated the effect of visual lifelogging data on the memory of users with no memory impairments. It has been shown that lifelogging images facilitate the ability to connect to the past [SFA$^+$07] and enhance the recall performance through end-of-day review sessions [FBB11]. However, a major disadvantage of lifelogging technologies, in general, is the sheer volume of captured data. In the case of lifelogging cameras, taking images every 30 seconds for about 12 hours operating time results in approximately 1,500 images per day. This exceeds the capabilities of users to review them all on a regular basis. Not only does this data contain many indistinct images; recording many blurred, dark or meaningless images additionally complicates the review process unnecessarily so that the envisaged beneficial effects on the human memory are weakened. From the users point of view, it would be desirable to filter out all irrelevant images and present the relevant images in a way that benefits the recall of episodic memories.

While the thumbnail view of file managers, image collages or comic-based presentations [Gir03, UFGB99, ĆGC07] already present a good overview about larger amounts of images, they do not show the images in full resolution which may weaken their effects as memory cues due to missing details. A

[1] Microsoft SenseCam Website: http://research.microsoft.com/en-us/um/cambridge/projects/sensecam/ (last accessed on October 10, 2015)

[2] NarrativeClip Website: http://getnarrative.com/ (last accessed on October 10, 2015)

[3] Website of the RECALL project: http://recall-fet.eu/ (last accessed on October 10, 2015)

slideshow of images is the most obvious approach to present images in their full resolution. Lee et al. presented relevant cues selected by caregivers in the form of a slideshow to memory-impaired patients and found that slideshows enabled them to think deeply about cues to trigger memory recollection [LD08].

For that reason, we investigate video summaries (i.e. slideshows of just the relevant images) that are designed to support the episodic recall. While prior work has already investigated video summaries created by caregivers to support memory-impaired people, we want to set a first milestone for the *automatic creation* of video summaries designed to support people with *no memory impairments* to recall episodic memories. For this purpose, we follow a user-centric approach and firstly conduct a user study to elicit requirements for video summaries that are designed to support people with no memory impairments to recall episodic memories. Next, we evaluate these requirements and compare a manual implementation of these requirements with review methods that present the entire lifelogging image set (i.e. timelapse and reviewing images manually) by measuring the effect on recall and the usability to gain further insights. Especially the usability may play an important role when it comes to elevating our concept to a mainstream technology.

1.1. Contributions of this work

In this work, we aim to develop a software that creates video summaries of lifelogging images and associated context data to support the episodic recall. This presumes a study in which we elicit requirements for such a video and gather empirical data to support the development. We use said requirements to implement the software and evaluate the effect of video summaries on the memory of participants.

Contributions of this work are hence

C-1: A set of requirements for video summaries designed to support the recall of episodic memories.

C-2: A comparison of video summaries to non-summarizing review methods (i.e. timelapse and reviewing all images manually) that reveals *(i)* no statistically significant difference in the effect on episodic recall and *(ii)* a better perceived usability.

C-3: A software implementation of these requirements.

1.2. Outline

After we introduced and motivated the topic of this work in this chapter, we discuss previous research on lifelogging summaries in chapter 2. In chapter 3, we describe a five-week study to elicit and evaluate requirements while chapter 4 present the result of the study. In chapter 5, we describe the development of the video summary creation software based on said requirements, which we then evaluate in chapter 6. This work is then summed up with a conclusion, discussion and directions for future work in chapter 7.

2. Background and Related Work

The research done in this thesis is based on general knowledge and former research in the field of lifelogging, psychology and image processing. To understand how to improve people's memory with lifelogging technologies, we first discuss the psychological background of memory and how to improve it. We will then present related work in the field of lifelogging including lifelogging systems and approaches on manage and presenting lifelogging data. This chapter will then be closed up with a summary and discussion.

2.1. Psychological Background: Memory

Information we receive from our environment are first processed by a series of sensory memory systems before they are passed to our short-term memory. Depending on the importance and quantity of rehearsal, information are then passed to the long-term memory from which memories can be drawn on even after years or the lifetime long. This information-processing approach is described by three stages of memory: (i) *encoding* through the sensory memory systems, (ii) *storing* in the long-term memory and (iii) finally the *retrieval* of memories later on [BEAA09].

According to Squire [Squ92], the long-term memory can be classified into two categories: the implicit (or nondeclarative) memory and the explicit (or declarative) memory [Figure 2.1]. The implicit memory refers to motor skills which reflect in performance rather than through remembering; examples are riding a bike, walking or playing the piano. In comparison, the explicit memory refers to knowledge that we can describe or reflect on. Endel Tulving [Tul72] proposed a distinction of explicit memory into two categories: the *episodic memory* that refers to events of one's own life such as appointments or holidays; and the *semantic memory* that refers to facts or information about the world, such as the color of a lemon [BEAA09].

In our work, we aim to use technologies to augment the aforementioned episodic memory of people. To understand how to do this, we have to first understand the limitations of our memory and how we can compensate this.

Schacter presented the misdeeds of our memory and classified them into seven basic "sins" of which three involves different types of forgetting: transience (gradual loss of memory over time), absent-mindedness (lack of attention while encoding information) and blocking (interference of similar information retrieved) [Sch99, CJ10]. Confirming the transience, Loveday *et al.* found that the failure to recall events of a day increased radically after a 5-day delay [LC11]. Anderson *et al.* presented the concept of retrieval-induced forgetting (RIF), which conforms to the blocking sin [ABB94]. Recalling specific memories can often lead to forgetting similar memories that competed for retrieval in the

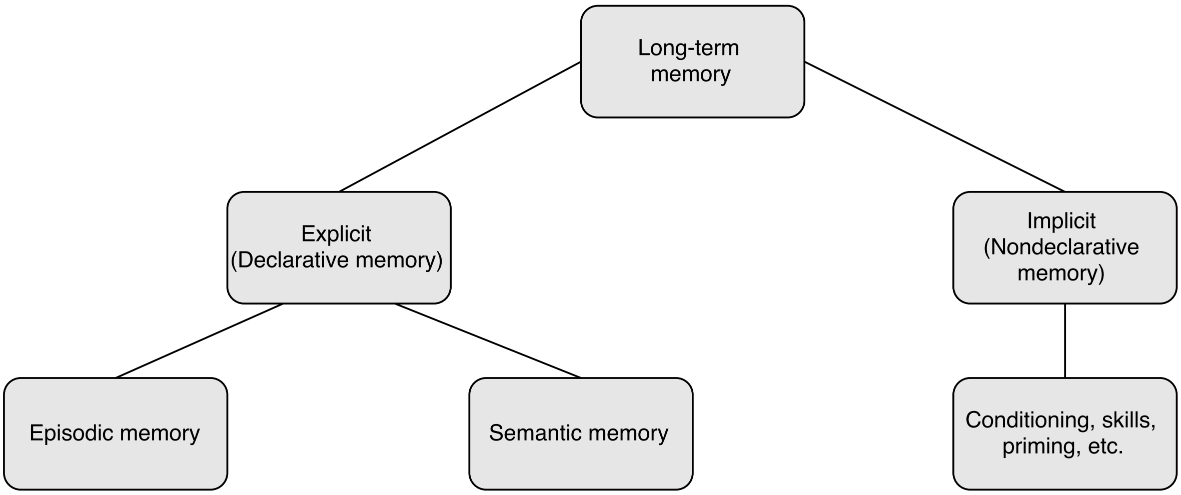

Long-term memory
Explicit (Declarative memory)
Implicit (Nondeclarative memory)
Episodic memory
Semantic memory
Conditioning, skills, priming, etc.

Figure 2.1.: Components of long-term memory as proposed by Squire [Squ92].

long-term memory. This was shown in Anderson *et al.*'s study in which participants practiced pairs of words (category – item) which lead to them forgetting the unpracticed ones. However, Migueles *et al.* presented a solution to this in their work, in which they explained how the usage of a so-called *script knowledge* of the daily routine can avoid effects of RIF and improves the recall of daily events [MGB12]. According to [BEAA09], "*script knowledge is a type of schema relating to the typical sequences of events in various common situations (e.g. going to a restaurant)*".

Tulving *et al.* introduces two states of memories in terms of recalling: *Accessible* memories are stored (and still available) memories that can be retrieved at any given point in time, whereas the *availability* indicates whether a memory was stored or not [TT73a]. Often, we have memories that are available but not accessible which often leads to the "on the tip of the tongue" experience. One example are inaccessible memories due to the consequences of RIF.

To help people recalling inaccessible memories, we need to provide memory triggers ("cues") as a memory aid. A cue is a related information to the memory one wants to recall that helps him to access the memory (e.g. a hint). According to the *encoding specificity principle* [TT73b], cues that are available at retrieval are more effective when they are similar to the condition that was present at encoding; or to say it in other words, cues are most effective when the information and context at encoding is also present on retrieval. Baddeley *et al.* have shown four types of memory that are retrieved best with context: environmental context-dependent memory, state-dependent memory (e.g. being drunk), mood-dependent memory and cognitive context-dependent memory [BEAA09].

2.2. Lifelogging

Data captured by lifelogging technologies has shown to be effective cues to help people recalling inaccessible memories [FBB11, SFA^{+}07]. Lifelogging devices store all kind of data that people encounter in their daily life: images, audio, videos, scanned copies of all sort of documents and further context information such as the location, appointments or even more adventurous information such as biometric data [SFR+13] or computer usage [CH02]. In 1945, Vennevar Bush had the first vision of such a lifelogging system that he described with "*a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility*" [Bus45]. About half a century later, Gemmell *et al.* from Microsoft Research finally fulfilled this vision with their project called MyLifeBits [GBL+02, GBL06]. In this project, one researcher captured all sort of documents, recorded audio, video and images, television and radio as well as transcripts of different communication channels during multiple years of his life. All this data is organized and can be accessed through a software that offers indexing and search functionality [GAL05, GLB03].

Similar memory prosthesis systems such as iRemember [VSB06], VAM [FO00] or Forget-Me-Not [LF94] capture data with the aim to support users in everyday tasks such as finding lost documents, recalling somebody's name by their face or recover information from past conversations. While aforementioned work aim to support the semantic memory, many researchers also looked into augmenting the episodic memory.

Previous work investigated the effect of images taken with lifelog cameras on the episodic memory. Results has shown that end-of-day review enhanced performance relative to no review [FBB11], and facilitate the ability to connect to the past [SFA+07]. Based on this, different methods to support episodic memory impaired patients to reminisce has been researched [LD07, KHBS10, BBK+11, LD08, YKK+09, WBH+15]. Similarly, research has also been done on supporting people with a healthy episodic and autobiographical memory. For example, Czerwinski *et al.* developed a system to help people return to their computer work faster by reminding them what has been done before leaving. Other systems augments the episodic memory in order to enable users to reflect on their lives [PCS+12].

Outside of the research, there is a huge amount of services that follows the initial idea of lifelogging. Chronos[1], Moves[2], Optimized[3], Argus[4], and the Apple Health App[5] are examples for applications that captures context data in order to help users to optimize their day. Not only do they capture data through mobile phone sensors, but can also be extended through activity trackers such as Fitbit[6] or Jawbone[7]. SAGA[8] further focuses on functionality to share captured data with friends over various platforms. Applications such as DayOne[9], Narrato[10] or Momento[11] place their focus on creating, managing and sharing journals of users day.

[1] https://www.getchronos.com/ (last accessed on October 10, 2015)
[2] https://www.moves-app.com/ (last accessed on October 10, 2015)
[3] http://optimized-app.com/ (last accessed on October 10, 2015)
[4] http://www.azumio.com/s/argus/index.html (last accessed on October 10, 2015)
[5] https://www.apple.com/ios/health/ (last accessed on October 10, 2015)
[6] http://www.fitbit.com/de (last accessed on October 10, 2015)
[7] https://jawbone.com/ (last accessed on October 10, 2015)
[8] http://www.getsaga.com/ (last accessed on October 10, 2015)
[9] http://dayoneapp.com/ (last accessed on October 10, 2015)
[10] https://www.narrato.co/ (last accessed on October 10, 2015)
[11] http://www.momentoapp.com/ (last accessed on October 10, 2015)

2.3. Managing Lifelogging Data

It requires a huge amount of effort for humans to manage the sheer volume of about 1,500 lifelogging images per day. Previous research presented techniques to ease this management, such as segmentation into main activities, summaries and detection of key moments.

Many segmentation approaches are based on MPEG-7 descriptors [Chi02, DS08a] which describe different characteristics of images through histograms. Based on SenseCam images and a user-annotated segmentation as ground truth, Doherty *et al.* found a solution based on peaks in dissimilarities between images [DS08a]. Other work approached the segmentation by using context information, such as a combination of bluetooth [BLD$^+$07], audio recordings [DSLE07], the GPS location [KBH$^+$14, CJG11, BLD$^+$07] or computer activities [CJG11]. Even activity recognition could be used to segment the lifelog data. Doherty *et al.* presented visual lifelog classifier that detects 27 different lifestyle activities based on MPEG-7 descriptors [DCC$^+$11]. Further approaches for activity recognition are based on object-hand interactions [FFR11], accelerometer [LKK11] or a combination of camera, microphone and accelerometer [BP$^+$06].

Segmentation of daily lifelog data is a first step to a more manageable data set, however, the result is still about 20 events per day [DS08a] consisting of 80-100 images each [BDSO08]. This is still a sheer amount of data which can be more condensed by representing those events through key images. One trivial approach is to just select the image in the middle of the event in question. Although that may work for frames in a video shot [SB06], this approach can be problematic with a set of lifelogging images that possibly contain bad images due to capture device limitations. Hence, different approaches consider these limitations and select key images based on the image quality [DBS$^+$08, BDSO08]. Cooper *et al.* presented another approach aimed to obtain distinct key images by selecting them based on their similarity to images in the same event and a dissimilarity to images of all other events [CF05].

Previous work presented methods that goes one step farther and use sensors to detect key images. Sas *et al.* for example used biometric data to filter images based on arousal and found out that high arousal images support richer recall of episodic memories than low arousal ones with over 50% improvement [SFR$^+$13]. Lee *et al.* used recognition of nearness to hands, gaze, and frequency of occurrence to discover important people and objects [LGG12], while Blum *et al.* approached this problem with detecting changes in activity [BP$^+$06]. Pärkkä *et al.* even used an extensive set of sensors containing amongst other audio, EKG, heart rate, pulse or skin temperature to perform activity classification with an accuracy of up to 86% [PEK$^+$06]. Doherty *et al.* relinquished additional sensors and presented an approach to develop automatic classifiers for visual lifelogs based on MPEG-7 Histograms and achieved an accuracy of 65% [DCC$^+$11]. Other related work also consider the novelty to detect key moments [DS08b, ASC11].

2.4. Presentation of Lifelogging Data

Previous work on the presentation of lifelogging data can be divided into two sub-categories: *(i)* presentation techniques that show images of the entire lifelogging image set, and *(ii)* presentation techniques that show only a past of the lifelogging image set.

Timelapses belong to the first category and are used to observe changes in a matter of minutes that otherwise would happen over hours or days. Lindley *et al.* investigated this approach in a field trial with household members [LHR$^+$09] while Berry *et al.* used timelapse videos an episodic memory aid for participants with memory impairments [BKW$^+$07]. Lifelogging images usually show many similar information when the user is not moving. Hence, it makes sense to combine timelapses with adaptive fast-forward approaches to skim the day even faster. Adaptive fast-forward approaches adapt the playback velocity on different characteristics, such as information density [HHWH11], similarity measures [PJH05], present motion [PD$^+$04] and manually defined semantic rules [CLCC09].

Not only do nearly identical images add no value, but also raise the chance of missing important information due to the sheer amount of visual data shown. The idea is to use lifelog filtering mechanisms and instead just show the relevant information.

Related work has investigated comic-like layouts [Gir03, UFGB99, ĆGC07] as a presentation method for a smaller amount of lifelogging images. Chiu *et al.* adapted these approaches and optimized them for mobile devices by using a voronoi-based layout [CGL04]. Boreczky *et al.* further tried out these comic book presentations to navigate through videos [BGGU00], while Lee *et al.* developed an interactive photo browser based on novelty values [LSO$^+$08]. Previous research looked into slideshows as a video-based summary approach to help people with episodic memory impairment on recollection of significant experiences [LD08]. Here, Lee *et al.* explained that slideshows allow to think deeply about cues to trigger memory recollection.

2.5. Summary and Discussion

In this section, we introduced the psychological background of memory and focused especially on the episodic memory and its limitations. These limitations refer primarily to memories that become inaccessible after a period of time although they are still available in the episodic memory. We learned that useful cues to support the retrieval of these inaccessible memories are featuring i.a. personal relevant content, novelty and a meaningful and recognizable context.

Lifelogging cameras have been shown to capture effective cues to support memory-impaired patients as well as people with no memory impairments. However, this presumes either a review of a sheer volume of captured images or the support of caregivers who prepare a presentation of lifelogging images that enables an effective recall of episodic memories. The huge amount of effort to review and the lack of simplicity are discouraging the general population from using these images as a memory aid since people with a healthy episodic memory are not as reliant upon such memory augmentation technologies as e.g. patients with an Alzheimer's disease. Hence, we presented related work aiming at reducing the complexity by segmenting, detecting key images in segments and understanding activities in the vast amount of visual lifelogging data.

Further, we showed prior work on presenting huge amounts of images, such as (different variations of) timelapses, comic-based presentation approaches or image browsers. Unfortunately, they all have in common that they present the entire lifelogging image set or an ill-defined part of it instead of comprehensible units as suggested by Byrne *et al.* [BLJS08].

As opposed to the presented work, we are aiming to develop an automatic video summary creation system that recognizes relevant image cues based on particular criteria and present them in the form of a video slideshow to enable an effective episodic recall. Since we want to focus on supporting the episodic memory of the general population, we follow a user-centric approach (as suggested in [SW10]) which presumes a requirements elicitation and evaluation phase before the software development is started.

3. User Study: Eliciting Requirements for Video Summaries

This chapter describes a five-week user study aimed to both gather requirements for video summaries designed to support the episodic recall, and to evaluate these requirements that are represented by a manual implementation created by the researcher. Gathered requirements will be used to inform the design of a video summary creation software that we present later in this work. It is well established that interviewees in requirements elicitation mostly express their desires unambiguously or overlooking important details while the interviewer might be biased towards his own ideas [LL12] which is why an evaluation is required. We further want to use the evaluation to compare the idea of a video summary to non-summarizing review methods (i.e. methods that present the entire lifelogging image set).

3.1. Research Questions

We designed the five-week user study to answer the research questions shown below. These address requirements on the video itself, an evaluation on the impact of such a video on the recall, and the advantages and disadvantages towards non-summary approaches such as a timelapse or a manual review of the lifelogging images.

RQ-1: How do participants manually create daily lifelogging video summaries to support the episodic recall of events occurred in their past (i.e. approximately one week ago)?

RQ-1A: What kind of images do participants include in a video summary?

RQ-1B: What are the characteristics of such a video summary?

RQ-2: In comparison to non-summarizing review methods, how effective are video summaries to support the review of events occurred in the past?

RQ-2A: How does the review method impact the level of recall of reviewed events?

RQ-2B: How does the user experience differ across review methods?

3.2. Methodology

In this section we present all information that are required to replicate this study at a later time. This includes the design, our apparatus, the participants and our procedure. Additionally to the procedure, we present our own method to measure the recall performance.

3.2.1. Design

The study is composed of five subsequent workshop sessions with a one-week interval elapsed between every session [Figure 3.1]. Although we conducted the study in one go, it can be thematically divided into two parts: Session 1 and 2 cover the *requirements elicitation*, while session 3, 4 and 5 cover the *evaluation* of elicited requirements that are represented by a manual implementation created by the researcher.

All workshop sessions include semi-structured interviews which are audio recorded. The requirements elicitation includes one semi-structured interview and an observed task in which participants are instructed to ’think-aloud’. The evaluation is conducted as a controlled experiment in which we used a repeated-measures design to compare the review methods with each other. The independent variable thereby is the review method (video summary, timelapse and manual review) whose order was counterbalanced across participants. We had two dependent variables which are (*i*) the impact on recall represented by our recall performance measure presented below and the memory experience [LS15], and (*ii*) the usability which is measured by two additional questionnaires [LHS08, Gro88] that are also presented below.

Participants were issued with a lifelogging camera to record one full day prior to their next workshop session and were reminded to do this by a weekly reminder e-mail of the researcher. The images captured on these days were then stored by researchers for a week before being used in the workshop sessions to create a video summary or to use them in the evaluation. Loveday *et al.* reported that the failure to recall radically increases after five days [LC11]; by ensuring a gap of eight days between capturing and using the images we could assume that participants had typically forgotten much of what occurred during the events captured and so could test the effect of the images for reminiscence purposes.

3.2.2. Apparatus

In the course of the study, we used three questionnaires to assess participants impressions on three different characteristics of a review method. These questionnaires are from previous work and are valid and reliable according to their authors. We used the NASA-TLX questionnaire [Gro88] to assess the cognitive load that participants perceived while using a review method to recall a past day. The user experience questionnaire [LHS08] by Laugwitz *et al.* was used to analyze the perceived user experience of a review method. The so-called memory experience was assessed with the questionnaire from Luttechi and Sutin [LS15]. The memory experience describes e.g. the coherence between daily events, sensory details or the visual perspective which are difficult to observe in interviews. Lastly, we measured the recall performance to analyze how well a participant could recall a past day before and after using a review method. Since we couldn't find any previous work on assessing the performance of a full-day recall, we decided to develop our own process that we described below.

We issued participants with a 1st generation NarrativeClip[1] as a lifelogging camera to capture images of their day. The NarrativeClip is a small and wearable camera with the size of 36 x 36 x 9 mm and can be attached to parts of clothes using a clip mechanism. The camera captures one image every 30 seconds when operational; the camera can also be manually triggered by tapping it twice. This device generates 5-megapixel images and includes 8 gigabytes of storage – allowing complete capture of at least one day, which is also the average battery life. The NarrativeClip has no on/off-switch and can be disabled by simply covering the lens (i.a. put the device up-side-down on a surface or in the pocket).

Both researcher and participants created video summaries using Google's Picasa software[2]. Picasa is an image management application that provide functionality to view images in a grid (similar to the thumbnail view of most file managers), assign images to albums and to tag them and to create a slideshow video.

3.2.3. Procedure

The study took place from May 2015 to July 2015 at the Lancaster University. The study started with a briefing session, in which participants were instructed on using the lifelogging camera and on creating a 'Narrative account[3]' to activate the camera. Additionally, we explained privacy concerns and exchanged contact data for potential problems during the study. Participants had 2-3 days after the briefing meeting to use the lifelogging camera for own purposes and getting used to it. All six sessions and five recording days are shown in Figure 3.1.

Participants used the lifelogging camera to record the day prior to the workshop session. They bring back the camera on the next day and transferred captured images to the researcher's computer first. In case participants captured images they didn't want to share with us, we allow them to delete these at the beginning of the sessions. The sessions then continue as follows:

Session 1: In the form of a semi-structured interview, we focused on gathering requirements about the video content, such as which memories they consider as valuable and what cues would help them to recall certain memories. We further asked a series of questions about their experiences and attitude towards lifelogging, their envisaged use case for video summaries, privacy issues and their ambitions and life goals. Life goals are determined supported by a questionnaire from Roberts *et al.* [RR00].

[1] NarrativeClip Website: `http://getnarrative.com/` (last accessed on October 10, 2015)

[2] Google Picasa 3: `http://picasa.google.com` (last accessed on October 10, 2015)

[3] Participants used the NarrativeUploader software (`http://start.getnarrative.com`) on the researchers computer to sign up for a Narrative account using their own e-mail address and password. We requested them to remember their login credentials to access their captured images during the course of the study.

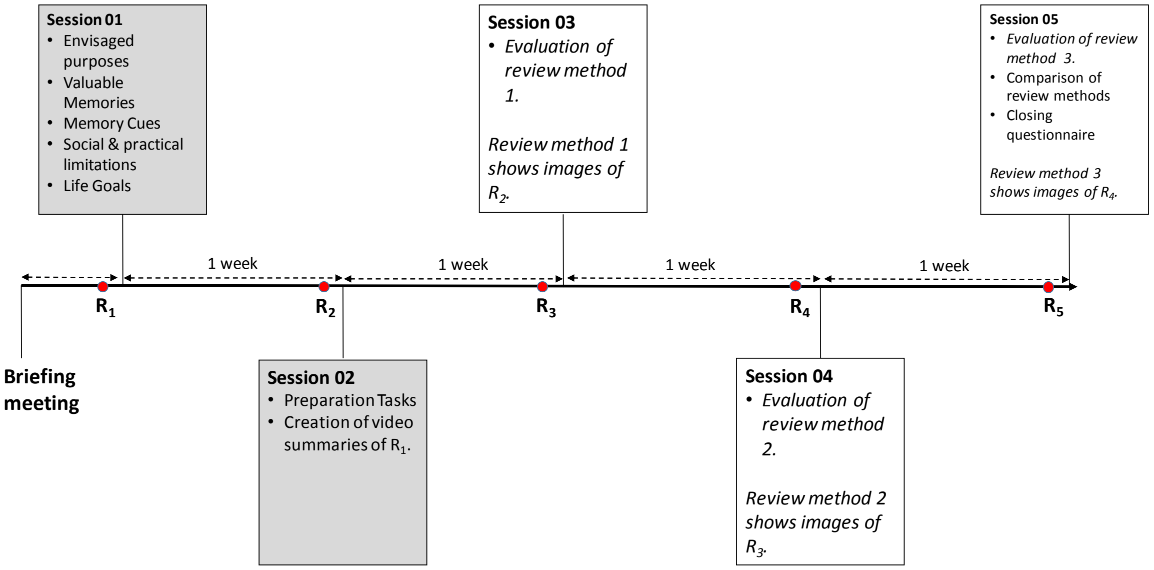

Session 01
• Envisaged purposes
• Valuable Memories
• Memory Cues
• Social & practical limitations
• Life Goals
Session 03
• Evaluation of review method 1.
Review method 1 shows images of R_2.
Session 05
• Evaluation of review method 3.
• Comparison of review methods
• Closing questionnaire
Review method 3 shows images of R_4.
1 week
1 week
1 week
1 week
R_1
R_2
R_3
R_4
R_5
Briefing meeting
Session 02
• Preparation Tasks
• Creation of video summaries of R_1.
Session 04
• Evaluation of review method 2.
Review method 2 shows images of R_3.

Figure 3.1.: The structure of the five-week study and topics of every session. Gray shaded boxes represent the sessions in which requirements were elicited and white boxes represent the evaluation sessions. The red points indicate the days in which participants used the camera to record images (R_i). Recordings always took place one day before a session.

Session 2: Using Google's Picasa software, participants generated a video summary based on the image set they captured eight days ago. This task allowed us to derive further requirements and to observe which images appeared to help participants to recall the past day.

In detail, we first prepare participants for the video creation task by letting them review their lifelogging images to recall the captured day. Next, we instructed participants to organize their images into clusters. We gave them no specific direction for creating these clusters, which allows them to come up with an own structure that help them to create the video summary later on. Finally, participants engaged in the creation of a video summary and were asked to 'think-aloud' as they completed this task. Participants were instructed to create video summaries that would help them to remember the most important things from the captured day. Upon completion, the created video was played back to the participant who was invited to explain the images that they had selected, the video composition, and how it helped them to recall the day.

Sessions 3, 4 and 5: Participants were asked to review images from the lifelogging capture day that had occurred eight days ago. Images were reviewed with the three review methods that we present below. Participants used a different review method in every session, whereas the order of the review methods are counterbalanced across participants. We rotated all images into their correct orientation before letting participants reviewing them. No further alteration was made.

- *Video Summary*: This is a personalized video summary that had been created by the researcher based on the requirements elicited in session 1 and 2. Requirements include selected types of cues and video characteristics that we will present in the next chapter. In short, a selected set of images ('the most relevant ones') are presented in the form of a slideshow following a chronological order. Each image is shown for the same duration that we found as optimal in the requirements elicitation phase[4].
- *Timelapse*: The timelapse is a much faster version of a slideshow that shows all captured images in a chronological order. Each image is shown for the same duration, whereas the duration of the whole timelapse video is about 2 minutes[5].
- *Manual Review*: To give participants the full effect of reviewing lifelogging images which has been shown to be an effective memory aid[SFA$^+$07, HWB$^+$06], we let participants browse their captured images in their own pace. Participants reviewed their captured images in Picasa's thumbnail view while a mouse enabled them to scroll through the thumbnail images and to click specific images to view larger versions at a higher resolution. Participants had no time restriction in this review approach. However, we measured the time they spent on completing the manual review.

After participants reviewed their images, we asked them to fill out three questionnaires: The NASA-TLX questionnaire for assessing the perceived cognitive load, the user experience questionnaire and the memory experience questionnaire. Additionally, we assessed the recall performance for recalling the recorded day each before and after the review of images. The method for this is presented below.

At the end of the study, participants filled out a closing questionnaire that revisited memory cues and envisaged use cases from the first session and covers demographic data about participants.

3.2.4. Recall Evaluation

The Recall Evaluation Process is part of the evaluation process for the approaches as described above. The aim of the Recall Evaluation Process is to determine how good participants are able to recall a past day either with and without one of the approaches. This process is based on the memory probe method described in [BEAA09, p. 141]. In a nutshell, the memory probe method is about giving subjects a cue and letting them recall anything related to this cue. The degree of details is then used as an indicator for recall performance. Sas *et al.* used a more concrete approach in their study [SFR$^+$13] to test the recall performance of subjects. Here, they asked participants for the *(i)* event, *(ii)* thoughts, *(iii)* emotions, *(iv)* place and *(v)* time after showing them a cue and rated the answers either with zero or one point depending on whether the aspect was stated or not.

In our study, we used a similar approach to the one used in the study from Sas *et al.* and extended it to evaluate the recall performance for a full day instead of just one activity. We asked participants to recall their day and telling us the three most important events for them on that day. We probed

[4] The duration is 3 seconds. We will present more details in the next chapter

[5] 2 minutes is the reportedly preferred length for a video summary.

for additional events if they were able to tell us at least three. Similar to Sas *et al.*'s approach, we prompted participants for details D_i for these three most important events. These details consists of the *(i)* time, *(ii)* place, *(iii)* thoughts and *(iv)* emotions associated with the event, *(v)* what happened during that event, and what happened *(vi)* before and *(vii)* after the event.

We rated given answers to the details D_i with a score of either 0, 1 or 2 points. An answer is rated with 2 points when the answer is complete, 1 point when it is incomplete. Examples for incomplete answers are *in the afternoon* instead of an exact time or *somewhere on the campus* instead of the exact place. Participants received 0 points for a detail if they can't give an answer to it. We subtracted the score by 0.5 points if participants hesitated while answering or inferred the answer from their daily routine or other activities. This was clarified by questioning participants.

For every of the seven details D_i, we then calculated the average detail score using all three activities from that day and added these scores to calculate the *average recall strength* (ARS) for that day. Hence, the formula for calculating the average recall strength looks as follows:

$$ARS = \sum_{i}^{Details} \frac{\sum_{j}^{Events} Score_{i,j}}{|Events|} * 0.14 \tag{3.1}$$

where

$$Score_{i,j} = D_{i,j} - P_{i,j}. \tag{3.2}$$

$P_{i,j}$ represents the penalty of 0.5 points for hesitation or inferring while $D_{i,j}$ represents the score of whether an answer was given and its completeness. Since we have 7 details and hence 14 is the highest score possible for the ARS, we decided to divide ARS by 0.14 to map the score range to 0 for the lowest score and 100 for the highest score. To finally obtain the Recall Performance Score RPS_d for the day in question we multiply the average recall strength ARS_d with the number of activities AR_d participants could recall for a day d.

$$RPS_d = ARS_d * AR_d. \tag{3.3}$$

We use the Recall Performance Score RPS_d in the remaining work as a measure for the success of recalling a past day d.

To investigate the improvement in recall after reviewing images with one of the three approaches, we let participants recall their day without any cue other than the date of the day they recorded eight days ago. We then let participants review their day with one approach and repeated the process in a second run. The improvement is equal to the difference between RPS_d before review and after review. In the second run, we focused on things that participants couldn't recall before the review. We also paid attention to errors that participants made during the recall before the review. Errors were discovered either by participants admitting that they recalled wrong information and by re-listening the recordings of this evaluation interview.

3.2.5. Participants

We had 16 participants (14 students; 2 staff) from the Lancaster University. Participants were between 18 and 39 years old (M=24.6; SD=5.4) of which 6 are female. Of these, 8 participants had reportedly never used lifelogging technologies while 2 participants reported minor experiences in using lifelogging technologies (i.a. short trials). Further investigations revealed that these lifelogging technologies are diaries, scrapbooks or applications to track the usage on computers and mobile phones.

Participants were recruited via a mailing list for studies at the Lancaster University. We rewarded the completion of the study with £50 and a copy of all their collected lifelogging images.

3.3. Privacy

Wearing a lifelogging camera through the whole day and sharing captured images with researchers can be a big privacy intrusion for many people. This effect is enhanced by many of our questions that addresses personal information such as interests, activities done on a day or information about friends and family. Hence, we designed the study to protect the participants privacy as good as possible.

The most vulnerable assets to protect are the images collected by participants. Fortunately, the NarrativeClip provides us with a security mechanism to protect all collected images with an account. Collected images could only be accessed when the participant logs into the account linked to the NarrativeClip. In case the NarrativeClip gets lost or stolen, pictures are automatically cleared from the NarrativeClip after linking the account to another NarrativeClip.

Participants are always the first to see their collected images and have the opportunity to delete images that they don't want to share with us. Hence, we asked them at the beginning of every session whether there are any images they want to delete or not.

To avoid participants to record inappropriate things or other people that are uncomfortable with it, we instructed them to disable the NarrativeClip when they are in situations where one has a reasonable expectation of privacy (e.g. restrooms, changing rooms or at a bank/ATM) or people around them request it. As described in the apparatus section, the NarrativeClip can be easily disabled by just covering the lens.

All privacy concerns are handled in the briefing meeting, where participants had to sign a consent form to confirm this. Each participant was assigned a unique number to identify the data set later.

3.4. Limitations

In a five-week study that requires 16 participants to return every week, it is almost naturally that some hurdles will occur. In our case, we had to postpone 9 appointments during the course of the study due to different reasons. For 8 participants, we had to postpone their appointment by 2 days due to technical problems (e.g. participants couldn't wear the device due to many meetings, or they forgot to wear it) while another 1 participant had to postpone his session by one day due to a spontaneous job interview. We further postponed all sessions of one participant by one full week due to sickness. However, all these postponements did not affect our defined minimum time gap between a recording day and the subsequent session (5 days) so that we can still assume that they had typically forgotten much of what occurred during the events captured.

Although we got provided with NarrativeClips from all participating universities of the RECALL project, we still had only 12 devices for 16 participants. Fortunately, we had enough participants who volunteered to return the device after a session and to collect it back two days before the next session.

We recorded most parts of the session using an audio recorder [6]. However, it would go beyond the scope of this master's thesis to transcribe and analyze all the recordings (over 30-40 hours of length), coding the answers all while analyzing the tremendous amount of captured images. Hence, we transcribed only parts of session 1 and took extensive written notes during session 2-5. We re-listened parts of the audio recordings of session 2-5 to extract valid citations from the interviews.

3.5. Summary and Discussion

In this section, we presented a five-week study with the aim to elicit requirements for a video summary designed to support the episodic recall of a past day. While requirements are elicited in the first two sessions with a combination of traditional and ethnographical elicitation techniques, the subsequent three sessions aim to evaluate those through a manual implementation. The evaluation allows us to validate the elicited requirements as it is known that interviewees tend to express their desires ambiguously while the interviewer might be biased towards his own ideas [LL12]. Moreover, we used these evaluation sessions to compare a manual implementation of the elicited requirements with two non-summarizing review methods (i.e. timelapse and manual review) by measuring the effect on the episodic recall and by letting participants assess the usability.

Although we had to deal with two study-related limitations, none of them violated the conditions for the independent variables of our study. Thus, we believe that the internal validity is not affected.

[6] Smart Voice Recorder for Android:https://play.google.com/store/apps/details?id=com.andrwq.recorder (last accessed on October 10, 2015)

4. Analyzing the Requirements and Evaluation

In the last chapter, we presented our study that aims to elicit requirements and evaluate them afterwards. In this chapter, we present the study results and derive requirements from them to inform the design of the video summary creation software.

We start with analyzing the social and practical limitations of lifelogging cameras which will give us an overview about the acceptance of such devices. Next, we focus on envisaged use cases, valuable memories and desired memory cues to get an idea of what participants expect from a video summary. We will then analyze the participants' video creation process to understand how these videos should be created. Results of the evaluation explain the impact of video summaries on the episodic recall and how valuable video summaries are in comparison to non-summarizing review methods such as timelapses and reviewing images manually. This chapter will be summed up with a summary and implications for the development of the video summary creation system.

4.1. Dataset

We collected a vast amount of data during the five-week study. All 16 participants captured 80 days in total which resulted in 69,250 images. On average, this means that one participant captured 865.23 images per recording day ($SD = 418.88$; $min = 111$; $max = 1960$). Workshop sessions were distributed as evenly as possible over the week, which resulted in 4 participants capturing each on Monday and Tuesday, 3 participants each on Sunday and Tuesday, and 2 participants on Wednesday. Since sessions all took place during the week, we have at least one captured image set for every weekday except Saturday.

Figure 4.1 shows the average amount of images that one participant captured per hour on one recording day. The least amount of images was collected at 4 am and increases slowly until 7 am. We can observe a steady increase starting at 8 am which ends with a peak at 3 pm. During 3 pm and 4 pm, 77.1 images were captured on average per participant on one day. During this time period, most participants either did their revision or worked so that they were mostly comfortable to wear the lifelogging camera. During 4 pm to 5 pm, we can observe a sudden drop of 17 images on average per participant on one day. Based on interviews, we assume that most participants were either doing sports activities, met their friends or had appointments at this time period which makes it difficult to wear a lifelogging camera. At 6 pm, we can see the local peak of the afternoon with an amount of 66.44 captured images per participant on one day. The number of images starts decreasing after this peak.

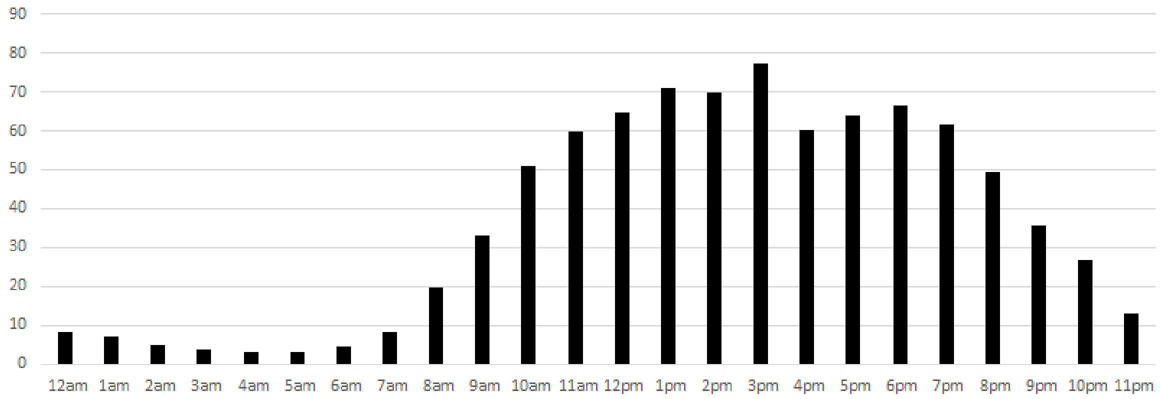


Figure 4.1.: Average amount of images taken by one participant per hour on one recording day. The X-Axis represents the time of the day. The Y-Axis represents the amount of captured images.

4.2. Social and Practical Limitations

We interviewed participants about their experiences on wearing the lifelogging camera and asked a series of questions about their attitude towards the idea of lifelogging. Experiences include the own behavior and the acceptance of lifelogging cameras in participants' social circle. This enables us to understand the implications on the completeness of image sets and how reactions and behavior may affect participants' memory of that day.

4.2.1. Reactions of the social environment

We asked participants about the reactions they got from people of their social circle.

Camera Conspicuousness: During the five-week study, 7 participants got reportedly approached by people of their social circle due to them noticing the lifelogging device. Another 2 reported that people *seemed* to have noticed the device but didn't show any reaction. The remaining 7 participants stated that people didn't notice the device by themselves until they were clarified about the situation on the participants initiative.

Acceptance of being recorded: 10 participants reported that their social circle agreed on being recorded without requesting any further explanations. 5 participants reported that they were initially requested to stop recording. However, the requesters changed their mind after they got cleared up about the study. Only 1 participant reported that one of his friends was uncomfortable to be recorded, so that the participant had to disable the camera in his friend's presence.

Behavior changes: The social circle of 2 participants reportedly got more conservative after noticing the lifelogging camera. People tried to avoid the camera by sitting beside the participant or even avoid them completely. Social circles of 9 participants didn't show any reaction and were comfortable with being recorded. 5 participants stated that they were not sure whether the behavior of people they met changed due to the lifelogging camera.

Additionally, 2 participants whose friends felt comfortable to be recorded even made some jokes about it ("*They smiled and waved into the camera. Friends made just a bit of a joke. They didn't feel uncomfortable*" - P4).

4.2.2. Behavior of Participants while wearing the NarrativeClip

We asked participants a series of question on whether wearing a lifelogging camera affected their behavior on the recorded day and whether this effect had a positive or negative influence on them.

Unintentional Recordings: 13 participants were reportedly afraid of recording things unintentionally while the remaining 3 had no worries about that. Counteracting the worries, we gave participants the opportunity to look through their images after they were transferred to the researcher's computer and delete all images that they didn't want to share with us. From all participants, only 7 participants deleted at least one image during all five sessions. From 80 image sets in total, there were only 11 sets from which images were deleted. 72.7% of the altered image sets are from the first or second session while only 27.8% are from session 3 and up. This trend might either indicate that participants got more used and gain more control over the camera over the first two sessions, or it might just indicate the laziness of participants who didn't invest enough effort on actually looking at their images to find unintentionally recorded images (e.g. scrolling through about 1,000 images in a matter of seconds). In total, 17 distinct images (134 images if including nearly identical and subsequent images) were deleted by participants.

Negative Effects: 5 participants reported that they *forced themselves to remember* to take off the camera when something intimate was done (e.g. going to the toilet). Another 3 participants reported an additional effort to obtain usable images ("*[..] also need to attach it to my clothes to get a good angle and pay attention to my hair to not cover it*" - P11). For 4 participants, the negative effect comes from people in their environment. Examples are uncomfortableness of friends (P10), regarding people's permission (P11, P16) or being made fun of (P3).

Positive Effects: 3 participants stated that they feel observed and hence were more productive on that day since they don't want to see pictures of them being lazy. P2, for example, explained, that "*the realization that you record your events helps you to stay [..] even more focused*". This behavior was also mentioned by many subjects in Chen and Jones' survey of over 400 general public participants [CJ12]. Further, 4 participants stated that the lifelogging camera was a conversation opener for them to get in contact with new people ("*It was definitely a conversation topic, so it was fun to wear*" - P4). Another 4 participants also saw the device as a kind of more practical camera in comparison to e.g. their smartphones. For 3 participants, it was a benefit that they have pictures of their day to review in case they miss anything important ("*I think if I just miss anything, I can go back and still see the images.*" - P7)

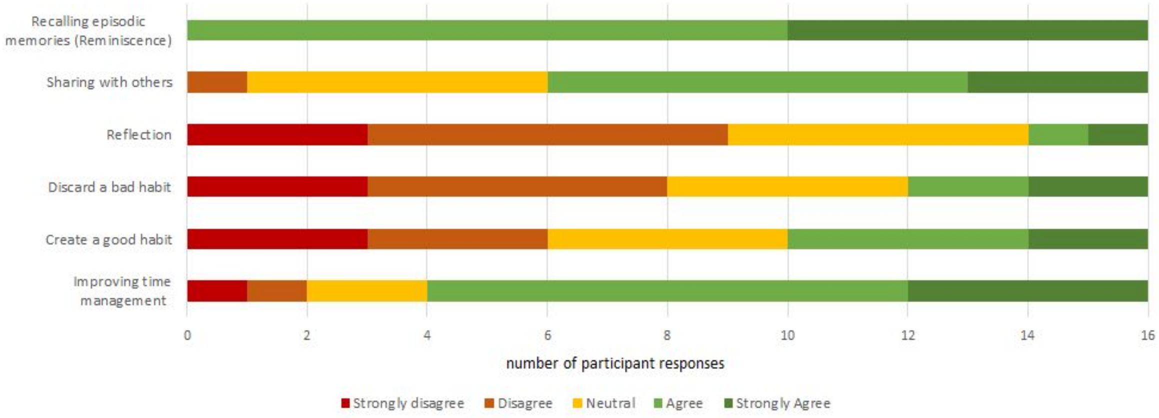


Figure 4.2.: Participants responses on the question "You want to achieve the following by using a video summary of your day". Answers were given on a Likert-Scale.

4.3. Video Requirements

In this section, we present the requirements we elicited through interviews, task observations and questionnaires. We start with promising purposes for video summaries and continue with memories that participants desire to preserve. With these in mind, we asked them about cues that they consider as useful to recall episodic memories and observed how suggested requirements were implemented in a practical video creation task.

4.3.1. Purposes of Video Summaries

Although our video summary is envisaged as a memory aid designed to support the episodic recall, we asked participants for further purposes for which they would consider using these video summaries. Promising purposes were collected in the first session and rated at the end of the fifth session (after all participants understood the concept of video summaries) with an option to suggest further purposes. Results are presented in Figure 4.2.

All participants agreed on the usage of the video summary for reminiscence purposes. 12 participants reported a desire to share video summaries with their friends and family. An additional question revealed that (of all 16 participants) 13 participants would share it with their friends, 10 participants with their family or life partner and 4 participants even considered to share it in social networks. 10 participants reported that they would use video summaries to reflect about their actions. Less than half of them indicated each an interest in using video summaries for self-improvement, such as creating/discarding a bad habit or improving the time management.

4.3.2. Valuable Memories and their Effect

Trying to convey every small detail is not only contra-productive ("*If we remembered everything, we should on most occasions be as ill off as if we remembered nothing.*" – James Williams, 1890), but also very boring and time-consuming for the viewer. Hence, we asked participants about memories they most wanted to preserve and for what reason they want to preserve these.

Positive Experiences: 15 participants identified that they want to preserve the positive experiences that happened in their life. Participants expressed positive experiences by terms such as "*good times*" (P12, P15), "*happy memories*" (P4, P7, P8, P13) and "*precious moments*" (P8). Examples for those experiences are e.g. traveling experiences (P1, P3, P8), special events such as birthdays or weddings (P5, P8) or time spent with friends of the family (P1, P9, P16). These memories are reportedly shaping their mood ("*When times are sad, you can remember those happy moments and it cheers you up.*" - P7) and making their life more fulfilling ("*Life is made of memories. If we don't keep them in any fashion, we are quite empty, right?*" - P9). Other participants simply realized that recalling positive experiences were enjoyable for them ("*It is nice to remember [...]*" - P16).

Social Encounters: 10 participants indicate a desire for preserving memories of social encounters. This includes experiences of "*hanging out with friends*" (P3, P4), getting to know new people ("*[..] Life-experience related to socializing and meeting new people, making new friends*" - P3) or times spent with an important circle of people ("*Time spent with family, friends and travel sister*" - P1). All participants agree that people are a very important factor in their life. P3 realized that people are redefining and inspiring him through their experiences and ideas in life ("*They sort of had me to redefine every time where I want to go and what I want to do. They inspire me. So that's why it gives me a new perspective.*" - P3) while P8 explained that people can bring him forward in his working life. Another participant stated a desire to preserve memories of others to draw on in times of loss or separation (*"Maybe you lose some people but you still have something to remind you of them*" - P1).

Self-Defining Memories: 7 participants stated a desire to preserve self-defining memories. These are experiences of overcoming demanding times (P2, P12) and difficult situations ("*Things that make you feel as if you push the limits of life a bit*" - P6), progressing in life (P2, P12, P6) and making mistakes (P13). A citation of P7 elaborates this the best: "*The memories of the way I grew up. Like, my struggles, my background. The way I developed and the way to where I am today*".

These memories help people to understand their personal identity ("*Bad memories that you overcame makes you stronger. You know.. if you've gone through a lot, especially a lot of hard times, it can make you a stronger person.*" - P12) and enable them to reflect back on past experiences to identify further goals in life ("*My memories show how I got to those life goals. What I've done. This is how far I gone. This is what I didn't do.*" - P2). The memory of overcoming hard times further affects the motivation for reaching further goals in life (P2).

Non-Episodic Memories: Unexpectedly, only 3 participants mentioned semantic information such as knowledge gained from studies (P10), conferences (P1) or outcome of conversations (P11). Additionally, P11 thought that statistics about food consumption and sports activities might be useful to reach a healthier lifestyle.

4.3.3. Desired Memory Cues

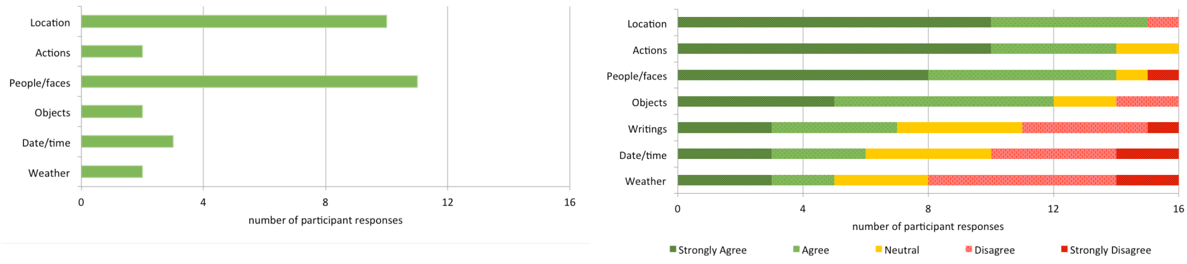
After we have learned about memories that participants desired to preserve, we asked participants for cues that they consider to be useful for episodic memory recall. To converge closer to our intention of creating video summaries, we restricted these cues to be representable by images. We investigated this topic twice in our study: The first time was during the interview in the first session and the second time was after the fifth session in the form of a questionnaire. Covering this topic twice allows us to investigate whether participants changed their mind based on experiences that they gained during the study by working with lifelogging media.

From the interview, we coded the responses into categories and present them in Figure 4.3a. We can see that the majority of participants regarded images of people (9 participants) and the location (12 participants) as most useful cues. Interestingly, 8 out of the 12 participants that regarded locational cues as useful are also regarding cues featuring people as useful. P12 explains that just the locational cue alone (in his case an image of a kitchen) isn't useful to recall a particular memory (since he is in the kitchen every day). Instead, he needs additional context (such as people holding objects or laughing) to enable him to recall the specific memory based on a better understanding of that event. The context of an event is further improved by additional cues such as objects (suggested by 3 participants) and actions (4 participants). However, actions cannot be depicted due to the static nature of images and are hence inferred from depicted people, locations or objects. P11 explained that "*daily usage objects*" tell her the actions, such as her desk is telling her that she is working at the office. Contrary to our question, participants also suggested metadata such as date/time or the weather.

At the end of the fifth session, we investigated memory cues again after experiences in lifelogging media were gained. This time, we listed all promising cues that were identified in the first session in a questionnaire and let participants comment on the utility of each listed cue. The results presented in Figure 4.3b indicate a change in the usefulness of some of the cues. While people (14 participants) and locations (15 participants) were still considered as the most useful memory cues, we can observe a significant increase in the usefulness of actions (14 participants) and objects (12 participants). We suspect three factors that might be responsible for this change: *(i)* gained experiences over the course of the study demonstrated them the usefulness of actions and objects, *(ii)* a different type of investigation (interview vs. questionnaire) that let them rate answers instead of finding answers and *(iii)* the recall evaluation in which we asked them about their day and got responses in the form of actions (e.g. revising, playing football, doing the laundry).

In terms of the metadata that participants suggested in the first session, we got 7 participants rating annotations as useful. 6 participants regarded date/time as useful and 5 participants liked the idea of indicators about the weather.

We observed that the four cues that were the most considered as useful (people, location, action and objects) are in line with the results of Lee *et al.* [LD07] who investigated memory triggers for people with episodic memory impairment.



(a) Cues identified in session 1.

(b) Cues rated in session 5.

Figure 4.3.: Types of cues that participants consider as useful to support the episodic recall. The left figure presents the cues we identified in the interview during the first session. The right image presents the rating (rated after the fifth session) of the cues that were identified in the first session. Answers were rated on a Likert-Scale.

4.3.4. Image Cue Features

We analyzed images that participants included in their video summaries and coded them into the categories that were identified for the desired memory cues. We will describe features of the images that belong into these categories based on their content and participants' think-aloud explanations. Moreover, we will show some examples of these images.

People-based cues: 23.2% of all video images show people that are a relevant part of the shown event. As people are reportedly one of the most useful memory cues, people-featuring images were always included when they were available (except for nearly identical images of the same person). A closer analysis reveals that people-featuring images have to meet certain criteria to be included in the video summary. Understandably, only people that affected an event or are known to the participant were selected. However, this is not always the case for e.g. life partners or close friends that usually do specific events together. P1, for example, included only a picture of her fellow students with each holding an ice cream to remind herself of how they hang out together after a lecture. She decided against including a picture of her boyfriend since they have the same lecture schedule. The identical schedule combined with the fact that they live together enables her to infer that he must have been with her during that event. P15 made the same decision and did not explicitly look for a picture of his girlfriend to represent an event that they always do together. However, participants did not explicitly aim to exclude images of their life partners. In case there is an image that shows their life partners with other relevant people, then that was the one that got included into the video summary.

Location-based cues: 45.0% of all video images were included to represent a specific location. However, we have to remark that this also includes a set of images that represent a location change (i.e. image that were taken while walking from one location to another). We observed that information about the location can also be inferred by objects or persons such as colleagues that the viewer only meet at work.

An important criterion for selecting location representatives are remarkable buildings. Unique color structures, shapes, signs or even labels make it easier to recognize the location than

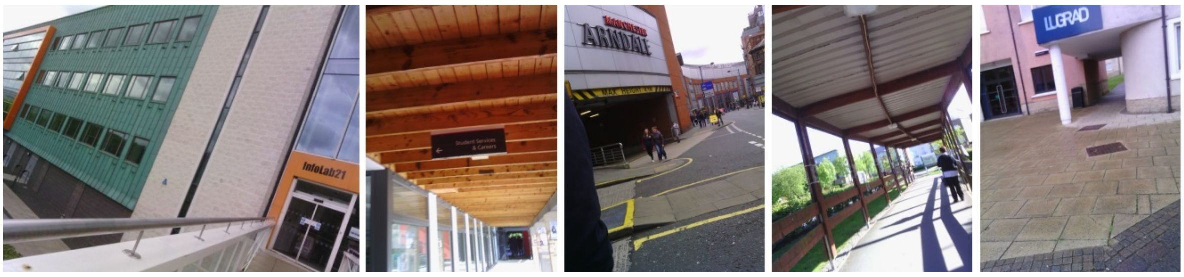


Figure 4.4.: Examples of useful location-based cues that allow an easy recognition of the location.

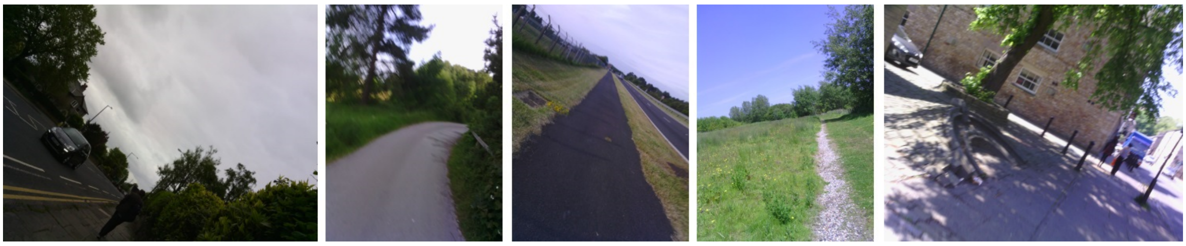


Figure 4.5.: Examples of less useful location-based cues. These images show mostly generic sceneries that could be easily confused with other locations. These images were not included in video summaries.

just generic scenery such as country lanes, forest tracks or unfavorable clippings of buildings. Examples of included images to represent locations are shown in Figure 4.4 while some bad examples (not included in the video summaries) are shown in Figure 4.5.

Object-based cues: 11.8% of all video images represent objects such as food, laundry, presents or dishes. While some of them were included to remind the viewer of the objects itself (e.g. lunch or flowers as a present), others were selected to represent an action such as cooking or revising. Examples are shown in Figure 4.6.

Action-based cues: 28.6% of all video images were included to represent specific actions. Actions are indirectly represented by static information such as people, locations or objects. Background knowledge is required to translate static information into memories of the action itself. Examples

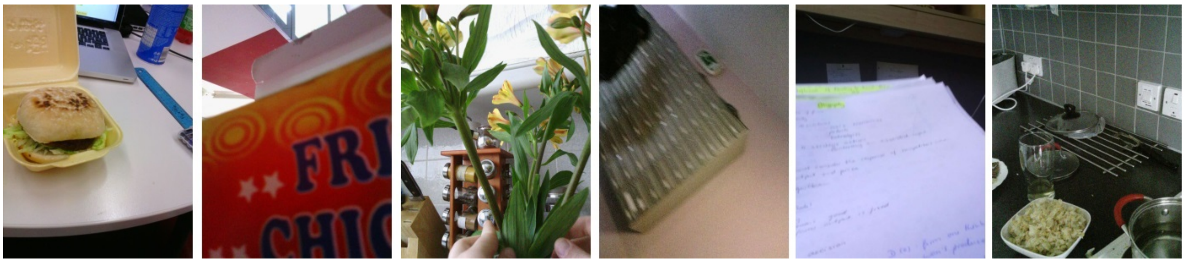


Figure 4.6.: Examples of object-based cues.

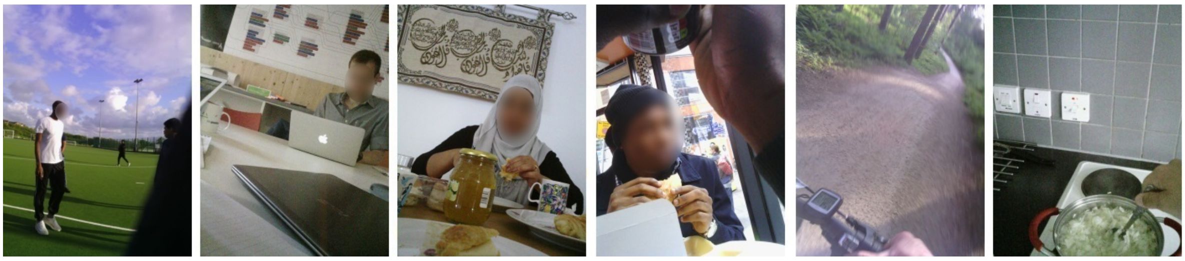


Figure 4.7.: Examples of action-based cues.

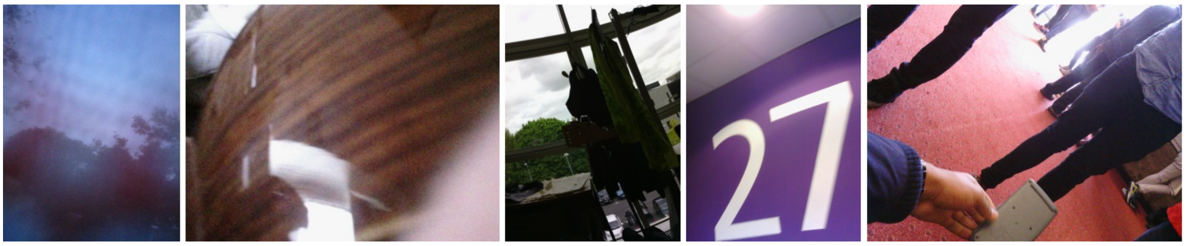


Figure 4.8.: Examples of images that requires background knowledge to understand.

are shown in Figure 4.7. The left image was selected by P7 who wanted to remind himself of playing football in his video summary. This can be inferred by seeing the football pitch and persons running around. The second image represents a working session which can be interpreted by seeing a person looking focused at the screen while using the keyboard.

Cues that depend on background knowledge: All participants are unanimous about the preference of clear images over defective ones (i.a. images that are difficult to understand due to blurriness, lens occlusion or unfavorable lighting conditions). However, participants had to resort to defective images due to limitations of the capturing device. In many cases, these images are only understandable with additional background knowledge of the participant. We showed some examples of included images in Figure 4.8. The left-most image seems to be a blurry image of some tree branches, which might not represent any information at all. However, understanding the blue, checked-like pattern in the foreground, P16 could recall that she was hanging up her laundry. The second image in Figure 4.8 shows a wooden pattern at the first glance, which reminds P09 of how he played his guitar since he can assign this information to his experiences. These images are still effective cues since participants are able to understand them and assign them to their original experience.

Moreover, there are also clear images that requires background knowledge to understand. These show specific objects that only the recording persons can understand due to memories and experiences that they connect with that specific object. An example is shown in the right-most image in Figure 4.8 that shows the recording person holding a calculator while standing on a red floor with many other people. With his background knowledge, he is reportedly able to translate this into a memory of him waiting in an exam hall.

4.3.5. Video Concept

We observed participants while they created a video summary. This enables us to gain further insights about their video concept and how that video concept helps them to improve the episodic recall. In the following, we present common concepts and characteristics that we could observe during the video creation process.

Short Video Duration: All participants would be willing to watch a video summary of no more than 2 minutes in length; 6 participants would accept 3 minutes, and further 3 participants would even accept 5 minutes in length. This indicates a desire to use video summaries as a quick review method to reminisce about a past day. The duration of the created video summaries confirms this; videos have 64.2 seconds in length on average (SD=35.9; min=39; max=188) and include 27.6 images on average (SD=21.03; min=6; max=81) while showing each image for 3 seconds (transition effects are not included).

Chronological Order: All video summaries present the images in a chronological order. This allows viewers to understand the interdependence between events and use their inferential process of retrieval to fill gaps in memory based on prior experience, logic, and goals; Baddeley *et al.* described this behavior as reconstructive memory [BEAA09, p. 180]. Reconstructive memories enable viewer to recall events that are not directly featured in the video.

This process becomes clearer with the following comment of P11 on an image that shows how she was walking to her office within the building she is working in: "*This image is redundant [. . .]. Of course I went to my office when I'm in the [workplace]*". In line with this comment, P12 watched a researcher-created video summary and noticed that since he wasn't shown revising in the library that day (his typical behavior at the time) but was instead in another room to revise, he must also have played football that day. The derivation of this activity (which he couldn't recall before) was made based on his background knowledge that this room was closer to the football pitch than the library is.

Distinct and interesting Images: Participants prefer to include cues that make the presented day different and special in comparison to other days. These cues include e.g. people that they normally wouldn't meet every day or places such as restaurants, meeting rooms or friends places that they wouldn't visit every day. P11 stated that if she would watch the video every day, she would only want to see the parts which are different from all the other days (novel events). According to [ASC11], it is generally accepted that novelty is very central in deciding whether to remember something or not.

Moreover, participants stated a desire for funny (P10), interesting (P6) and aesthetically pleasing images (P8) which have a greater likelihood of triggering past memories. P6 and P13 used the lifelogging camera to take pictures together with their friends and explicitly looked for these during the video creation since these were mostly funny pictures.

These findings are in line with the guideline from Byrne et al. for presentation and visualization of LifeLog content in which they suggested that cues should be "*enjoyable, rich and engaging but also meaningful*" [BLJS08].

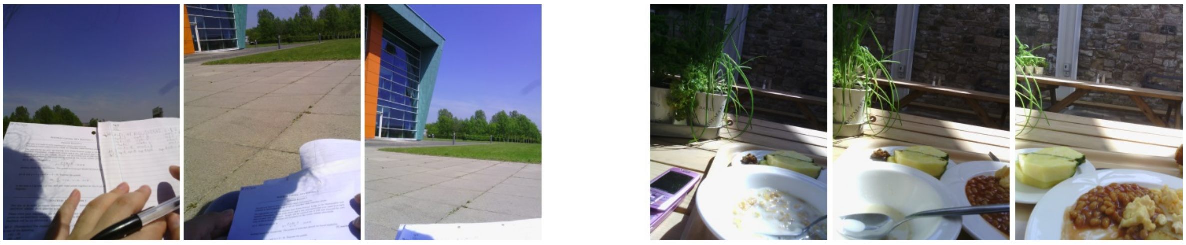


Figure 4.9.: Examples of images representing a static event and showing the most information at once out of a set of very similar images. This allows to see all information in one image instead of being spread over multiple images. In the two examples, the image in the middle was selected.

More images for periods with greater movement: Analyzing the images, we found out that 10 participants included more images of events with greater movement (e.g. walking to the library) than events in which they are at the same place for a longer period of time (e.g. revising in the library). Of all images selected for video creation, 32.5% shows only the route to the next location and hence are representing a movement exclusively.

Since periods with greater movement result in more different images, small gaps are more required to help to understand what happened after consecutive images. In comparison to that, events in which the recording person is not moving mostly results in similar images for a longer period of time. P8 also justified this with the pictures being different and nicer.

Display Time per Image: Participants (P3, P12, P14, P15) stated that they need enough time to understand an image and think further about it (i.a. to reconstruct further memories based on the inferential process). During the video creation process, the standard setting of 3 seconds was not altered by any participant. A shorter duration would be reportedly overwhelming for viewers since they would be interfered in their inferential process by too many incoming information which may lead to confusion (P14). A longer duration could lead to boredom since viewers have to wait (and therefore feeling like they wasting time) for the next image after understanding and thinking about the currently shown image.

Static event representatives with most information at once: From a set of images representing a static period (similar images), participants tended to select images that show the most information at once. For example, to represent a revision session captured by 164 images in front of a recognizable building, P5 invested effort to look for one that shows both her notes and the building. The majority of images either showed the building or the notes, but not both at once. Two examples of this behavior are shown in Figure 4.9. The image in the middle was the one selected while the adjacent two are from the same static event.

4.4. Evaluating Manual Implementations of Elicited Requirements

After the requirements elicitation phase, we created a video summary for each participant based on the requirements we elicited. This handcrafted video summary was then evaluated and compared to two non-summarizing review methods; the timelapse and reviewing the images manually. In this section, we describe the impact of our video summaries on the episodic recall and compare the video summary with the two non-summarizing review methods.

4.4.1. Impact on Recall

All participants agreed that our video summaries helped them to recall all major events of their past day (e.g. "*every major thing is captured in the video*" - P6). P9 was even surprised about the quality of our videos on summing up his day: "*I am a bit surprised with the quality of the images in terms of summing up the event [..]. As a memory cue, it's quite effective to sum up the things*". He further stated that the video gives him "*good memories about the guys that were playing [football] with*" him and emphasized that he likes that all important people are in the video. Further participants complimented the video summary on being "*very good*" (P2, P8). However, three participants criticized that the video summaries contain too many useless images. They pointed out especially the images that we included to represent events with greater movement (see above).

Video summaries reportedly provide viewer enough time to recognize, understand and think about a shown image. P9, for example, describes the video summary as a "*compressed review method that focuses on relevant information*" while, for P6, the video summary is an interesting, enjoyable and story telling review method. P3 likes the fact that he had enough time between the images to "*integrate what [he] just saw and use this time to recall other things*".

We asked participants for memories that were cued by the video summary and that they couldn't recall before seeing the video summary. We coded these memories which fell into the following categories: *(i)* actions, *(ii)* people, *(iii)* location, *(vi)* minor details (e.g. weather, time, order of events) and *(v)* objects. Figure 4.10 shows the amount of cued memories for each of the categories. We can see that participants are able to access especially memories of actions and people again after they saw the video summary. There were also 6 cases in which the video summary cued memories of participants being in a specific location or minor details such as the weather or the time at which something happened. Three participants noticed that they recalled something wrong before they saw the review of the day.

Without any cues other than the date of the day they recorded, participants could recall AR_d = 2.68 events (SD=1.08; min=1; max=6) on average. After watching the video summary, we observed an improvement of 1.38 events (SD=1.45; min=0; max=5) additional events on average which makes AR_d = 4.06 (SD=1.29; min=3; max=7) events in total. We could also observe an improvement in the average recall strength (ARS_d) which was assessed based on the information that participants were able to recall about the seven details D_i (as a reminder, these were *(i)* time, *(ii)* place, *(iii)* thoughts and *(iv)* emotions associated with the event, *(v)* what happened during that event, and what happened *(vi)* before and *(vii)* after the event). Before using the video summary, participants scored an average recall strength ARS_d of 66.03 (SD=15.56; min = 35.71; max = 97.62) which improved to 76.26 (SD=12.56; min =52.38; max = 97.62) after watching the video summary (the maximum achievable score is 100). Multiplying the average amount of events AR_d with the average recall strength ARS_d results in the recall performance score RPS_d that we present in table 4.1.

	RPS before cue	RPS after cue	RPS Imp.	ARS Imp.	AR Imp.
Video Summary	182.4 (86.4)	308.9 (103.6)	126.5 (128.9)	10.2 (15.7)	1.4 (1.5)
Timelapse	144.9 (95.6)	296.5 (143.7)	151.6 (141.2)	18.1 (26.0)	1.6 (1.5)
Manual Review	176.6 (121.1)	253.8 (117.6)	77.2 (65.0)	12.2 (20.1)	1.1 (1.1)

Table 4.1.: Recall performances measured with out recall performance evaluation described in chapter 3.2.4. The columns on the left describe the recall performance score before and after reviewing images. The columns on the right describe the improvements in the recall performance score (RPS), average recall strength (ARS) and the average amount of events recalled (AR).

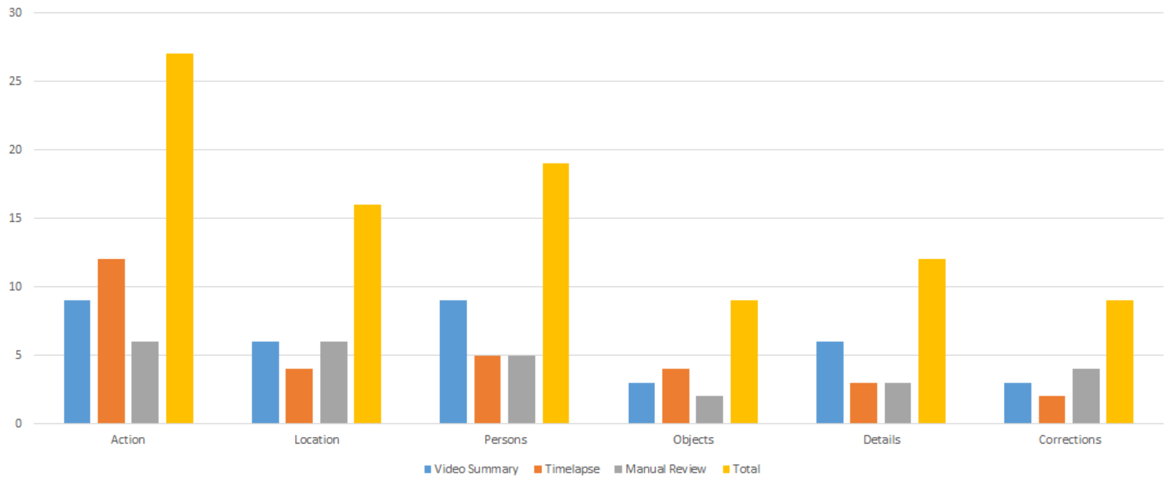


Figure 4.10.: Memory types that participants couldn't recall without a prior image review.Bars indicate the amount of memories recalled (all participants).

4.4.2. Comparison to Non-Summary Review Methods

We compared the video summary to the non-summarizing review methods (timelapse and manual review) based on two aspects: the impact on recall and the usability.

Comparing the Impact on Recall

In the following, we will first compare the recall performance improvements (RPS_d) and the results of the memory experience questionnaire.

Recall Performance Measure: Table 4.1 presents the improvement in our measure for the recall performance for all three review methods. The largest difference (RPS_d) between recall performance before and after a review is seen when using timelapse as the review method, followed by video summary review and finally manual review. However, a one-way repeated-measure ANOVA does not reveal any significant difference between the three review methods, $F(2,30) = 1.421, p = .257$.

Memory Experience Questionnaire: Using Luttechi and Sutin's [LS15] memory experience questionnaire we see a clear improvement in ratings for memory vividness, coherence, accessibility, time perspective, visual perspective, sharing, distancing and valence. In each of these cases video summary scores more highly than timelapse, which in turn scores more highly than manual review. By contrast, for the emotional intensity and sensory detail dimensions a different trend is seen – for these timelapse scores more highly than the video summary, which in turn scores more highly than manual review. However, despite this trend, a Friedman ANOVA does not reveal any significant differences except for the time perspective, $\chi^2(2) = 8.415, p = .015$. No other comparisons were significant.

Categories of Memories recalled: Figure 4.10 shows the amount of memories recalled for each memory category. The most interesting differences can be observed in the categories of people and details. Here we can see that the video summary helped participants to recall memories of people and details more than the other view methods. However, we can't be sure whether the video summary can be credited for this since this also depends on what happened on the recorded day.

Subjective Assessments and Comparisons

Although there is no significance difference in the impact on the episodic recall, participants still prefer the video summary due to subjective reasons.

User Experience Questionnaire: Laugwitz *et al.*'s questionnaire [LHS08] indicates that the video summary yielded a better user experience on every attribute than the timelapse review method, which in turn yielded a better user experience than the manual review method [Figure 4.2]. However, a one-way repeated-measures ANOVA revealed significant differences only in attractiveness ($F(2,30) = 9.938$, $p < .001$), stimulation ($F(2,30) = 9.168$, $p = .001$) and novelty ($F(2,30) = 4.103$, $p = .027$). For each of these attributes, Bonferroni post hoc tests revealed significant differences between the video summary and manual review. No other comparisons were significant.

Cognitive Load Questionnaire: Results from the NASA-TLX Questionnaire [Gro88] indicate that timelapse appears more cognitively demanding than manual review which in turn is considerably more demanding than the video summary [Figure 4.11]. A one-way repeated ANOVA reveals that there is a significant difference between the three review methods, $F(2,30) = 10.747, p < .05$. Bonferroni post hoc tests revealed a significant difference between the video summary and the timelapse, $CI_{.95} = -37.591$ (lower) -8.534 (upper), $p < .05$ and between the video summary and manual approach, $CI_{.95} = -34.333$ (lower) -5.917 (upper), $p < .05$. No other comparisons were significant.

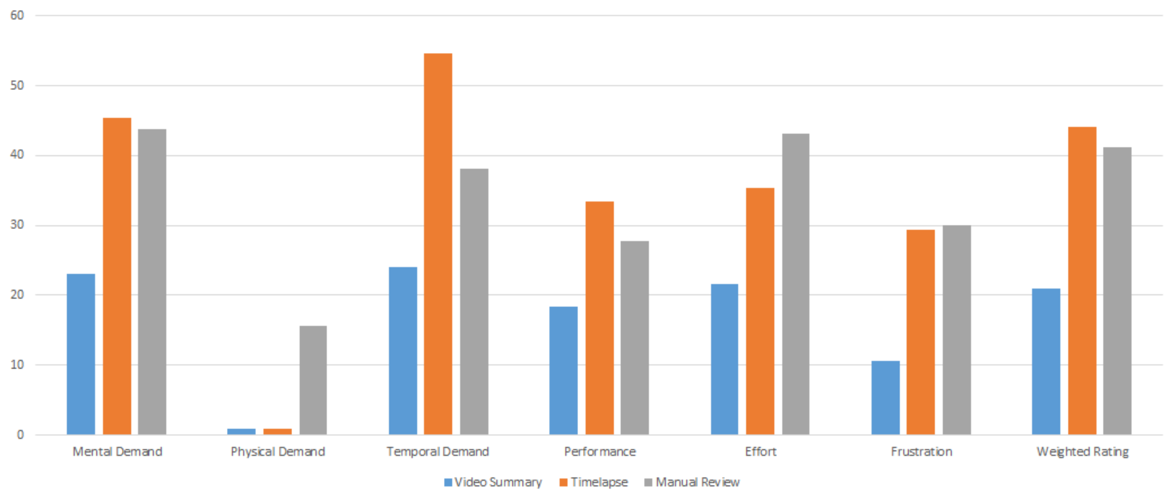


Figure 4.11.: Results of the NASA-TLX questionnaire that assesses the perceived cognitive load.

	Video Summary	Timelapse	Manual Reviewing
Attractiveness	1.8 (0.9)	1.1 (1.2)	0.3 (1.1)
Perspicuity	2.0 (0.6)	1.4 (1.1)	1.3 (1.1)
Efficiency	1.4 (0.7)	1.0 (1.1)	0.6 (1.1)
Dependability	1.1 (0.6)	0.6 (0.8)	0.5 (0.8)
Stimulation	1.5 (1.0)	0.9 (1.0)	0.3 (1.2)
Novelty	0.9 (0.9)	0.4 (1.3)	-0.3 (1.3)

Table 4.2.: Results of the user experience questionnaire from Laugwitz et al. [LHS08]. The values in brackets describe the standard deviation.

A subjective comparison to the non-summarizing review methods confirms these results. Participants compared the video summary with the two non-summary review methods and finally ranked all three methods by their usefulness as a memory aid for the episodic recall.

Comparing to the timelapse: Many participants (8) complained that the timelapse was too overwhelming (P11, P14, P3), confusing (P12), easy to miss out information due to the pace (P15, P16), and did not allow participants to think about or recognize what has been seen (P8, P10). In line with the results from the cognitive load questionnaire, this is what causes the high perception of cognitive load, especially the temporal demand and frustration. P14 even stated that he couldn't recognize or understand much since he moved a lot on that day which resulted in a very unsteady video. In comparison to the video summary, P14 also assumes that he would have recalled more events of the past day when he had more time to think about particular images instead of being overwhelmed with new information over and over.

The fast pace of the timelapse doesn't only overwhelm viewers but might also lead to actually missing important events. This was exactly the case for P16, who couldn't spot specific images in the timelapse that she took manually (with the double-tap functionality of the NarrativeClip).

After she reviewed her images using a file manager, she could find the images in question and confirmed that she indeed overlooked them in the timelapse. Besides the perception of a high cognitive load and situations in which images can be overlooked, participants also commented on the timelapse as boring due to the repetitiveness of images (during events with less movement) and due to many useless images (due to blurriness or lens occlusion).

However, P10 and P15 consider the presentation of all images as an advantage in comparison to the video summary since there is no risk of losing important images. In comparison to the video summary, presenting all images gives viewers a feeling for the duration of an activity (assuming they recorded the whole event) and feels more like a journey through the day (P6).

Comparing to manual reviewing: A manual image review isn't only more time consuming and exhausting (P3, P9), but it also doesn't provide the same point of view as the video summary or timelapse (P10: "*It doesn't feel like I'm the same person than in the video*"). Opening and closing the full view of an image further requires extra work which leads to a loss of focus and tiredness for a big amount of images (P3).

One disadvantage in comparison to a video summary or timelapse is the uncertainty of time required to review. During the study, we often had the suspicion that participants tried to hurry up to review all their images (which is about 865.23 on average) in a reasonable amount of time that they would be willing to spend in a real-world setting. On average, participants spent 153 seconds on reviewing images manually (SD=91.31; min=30; max=445). In contrast to hurrying up, they also tried to pay attention to the details of the images which is difficult to do due to the small thumbnail size. This conflict is reflected in the mental demand, temporal demand and especially the frustration in the cognitive load questionnaire.

The advantage of the manual review in comparison to the video summary and timelapse is the greater control ("*The power is in your hand.*" - P12) and the perceived level of detail due to the greater control (P9).

Ranking the review methods: To summarize all factors up, we asked participants for a final decision on which review method they would use as a memory aid to support their episodic recall. Participants ranked the review methods from the 1st place (as the most preferred) to the 3rd place (the least preferred). Figure 4.12 shows a clear preference for the video summary, with 10 participants considering the video summary as their first choice to support their episodic recall.

4.5. Life Goals

Life Goals are important indicators for memories that people want to keep. The kind of memories that one wants to remember differs between people with different life goals. For example, somebody who wants to live a healthy lifestyle pays more attention to e.g. healthy food and regular sports activities while somebody who wants to improve their social life requires to remember faces and names of persons they've met.

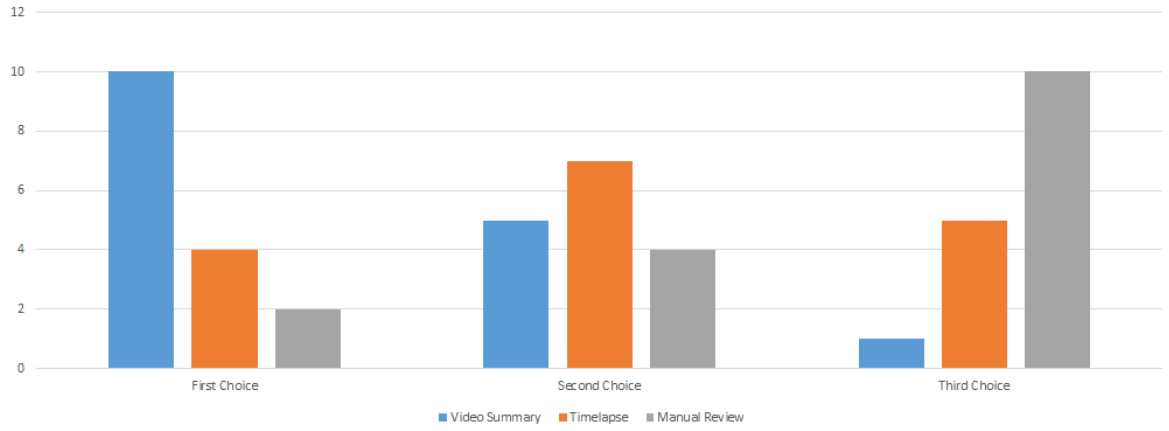


Figure 4.12.: Participants subjectively ranked the three approaches by which they think is the most useful episodic memory support for them.

We used a questionnaire from Roberts and Robins to determine Major Goal Clusters in their life [RR00]. Table 4.3 shows the result of this questionnaire, whereas emphasized numbers represent one of the top three life goals of a participant. We can see that all participants follow economic goals in their life (e.g. getting a well payed job), while 13 participants follow relationship goals (e.g. marriage). 12 participants followed hedonistic goals in their life (e.g. having fun).

Our analysis does not reveal any relationship between these life goals and the video creation behavior (i.a image cue selection). On the one hand, we have mostly very similar results as all participants followed economic goals while the majority of participants (12 and 13) followed hedonistic goals and relationship goals. On the other hand, we had 15 of 16 participants creating a video summary based on images of a rather normal and uninteresting day in which all were either revising or working. Hence, these images doesn't show us any sign that participants were following a life goal except the economic goals.

To investigate the relationship between selected pictures for video creation and participants interests and life goals, it might be better to conduct this study during a phase in which the participants are less focused on working towards one single goal which is the exam phase in this case. Then, we could have more opportunities to learn about what is interesting for participants and how they would involve these things into a video summary.

4.6. Summary and Discussion

In this chapter, we presented the results of the five-week study and provided answers to our research questions stated in section 3.1. We presented the requirements for a video summary designed to support reminiscence (RQ-1) and an evaluation of a manual implementation of these requirements (RQ-2).

	Economic	**Aesthetic**	**Social**	**Relationship**	**Political**	**Hedonistic**	**Religious**
P1	**16,08**	7,96	10,10	**11,81**	9,74	**11,95**	7,49
P2	**22,44**	4,80	8,63	**13,95**	**12,06**	9,82	6,84
P3	**21,20**	**13,50**	2,49	10,79	**12,08**	11,69	8,53
P4	**19,97**	3,60	3,26	**13,95**	5,78	**11,95**	8,53
P5	**14,96**	**9,31**	5,52	8,47	4,73	**9,21**	7,11
P6	**12,48**	**14,03**	8,35	9,62	5,18	**11,69**	6,17
P7	**19,85**	7,57	9,05	**13,18**	9,24	**10,58**	10,04
P8	**18,20**	4,30	6,04	**10,62**	7,35	**11,69**	6,09
P9	**15,03**	11,90	6,04	**13,12**	10,46	**11,95**	6,83
P10	**20,56**	7,20	9,98	**11,64**	8,31	**11,95**	6,96
P11	**19,48**	11,50	8,27	**13,95**	7,00	**11,95**	8,87
P12	**18,49**	6,60	10,61	**13,95**	10,02	**11,95**	10,71
P13	**14,86**	3,60	5,67	**9,01**	6,15	**7,17**	3,41
P14	**20,21**	**15,60**	8,27	**13,95**	9,78	11,95	11,13
P15	**17,44**	5,00	4,74	**11,98**	8,72	**11,34**	3,97
P16	**13,37**	4,80	5,94	**12,94**	3,51	8,95	**10,00**

Table 4.3.: Results of the questionnaire about major life goals from Roberts and Robins [RR00]. Values indicate the relevance of the respective major life goal to the participant. The three most important life goals are emphasized.

We learned that participants have a desire to preserve memories of positive experiences, social encounters and self-defining experiences to shape their mood, motivate themselves or reflect back on the past. To help to recall these memories, image cues featuring a combination of persons and locations, objects or action cues were regarded as the most useful. This has also been shown in other research work [LD07]. Additionally, people featured in images should be relevant to the shown event (i.e. not a person in the background) while locations should be presented by remarkable structures and environments. Optimally, images should be distinct and interesting (e.g. hobby-related, aesthetic or special) which is in line with findings in [BLJS08]. These findings allow us to regard RQ-1A as answered.

To further enhance the insights of participants' image selection process, we gathered the life goals and ambitions of participants using a questionnaire that assesses the relevance of seven major life goals. Since the majority of our participants (14 of 16) are students and were all preparing for the exam phase at the time of our study, we couldn't find any meaningful relationship to any life goal other than the economic goals (which is also the most important life goal for all participants).

Answers to RQ-1B were presented in the form of the video concept. Here we found out that cues have to be presented in a chronological order to enable the video summary to convey far more information than shown on the images. By presenting images in their chronological order, we activate the script plus associated background knowledge of the viewer [MGB12] which helps to reconstruct memories of events that are not directly featured in the videos. Moreover, this helped participants to understand impaired images due to the interdependence between events (daily routine: after the event a, I do event *b*). In line with this, Bower *et al.* found out that cues organized by the temporal order of the script knowledge are "far superior and more organized for scripted activities than when presented in random order" [BCM80].

Participants preferred to use video summaries as a fast way to reminisce over a day. Hence, they agreed on a maximum duration of 2 minutes and created video summaries that are even shorter. However, images should be displayed for about 3 seconds to enable viewers to fully recognize, understand and think about it. 3 seconds per image and a 2-minute time-limit means that there are at most 40 image slots to fill. To not waste any slots, participants deliberately look for images that contain the most amount of information out of similar ones. If possible, images that doesn't show any meaningful information (e.g. blurry images or images of the ceiling) were avoided.

Later in this chapter, we present answers to RQ-2 in the form of an evaluation and comparison. These revealed that although there is no significant improvement in the recall in comparison to the timelapse approach (RQ-2A), participants perceived summary videos as a much more positive approach in terms of user experience (RQ-2B) which is an important factor for the application as a mainstream technology. In general, our video summaries support reminiscence in a different way than review methods that show the entire lifeloggimg image set. While non-summarizing review methods convey every available information down to the last detail, video summaries show just an overview of the most relevant events and leaves room for the viewer to reflect on that. This allows the inferential process to reconstruct memories that are not even featured in the lifelogging image set and might lead to a better recall than when being overloaded with plenty of small and irrelevant details, such as it is the case with the timelapse review method. However, the efficiency of this method relies on the inferential process of the viewer.

There are some limitations in our study that were not avoidable due to the nature of studies on the memory. According to Loveday *et al.*, the failure to recall radically increases after five days [LC11]. We decided to let participants use the captured lifelogging images after eight days to ensure a gap of at least five days in case we have to reschedule the sessions. However, there are some captured days (13 of all 80 days) that participants could recall nearly perfectly while the remaining days are in line with the report of Loveday *et al.* This is due to days in which participants experienced something very emotional (e.g. P6 losing her identity card before her exam) or which was distinct to other days (e.g. P11 going out to dinner after work). Further, the results show that participants were afraid of recording things unintentionally. This may cause a feeling of uncertainty that in turn may lead to a stronger memory of the captured day. While the first case may be unavoidable, we attempted to diminish the feelings of uncertainty by a short probe phase of two days after the briefing session to let participants getting used to the study setting.

In total, these requirements and evaluation results gave us a clear picture of how to design a video summary to support the episodic recall. Additionally, we learned about the strengths and weaknesses of video summaries to further build on them.

5. Software Requirements and Implementation

In the last chapter, we presented the results of the requirements elicitation and summarized them at the end of the chapter. In this chapter, we will follow that up and describe the vision and the requirements for the video summary creation software based on last chapters summary and its implications.

Regarding the implementation, we will first present basic technologies and frameworks that we used to interpret the images. Afterwards, we present the design and implementation of the software and close up this chapter with a summary.

5.1. Requirements and Vision

Based on the results presented in the last chapter, we summarized the following requirements for our video summary creation software:

1. **The duration of the video summary must not exceed 2 minutes.** Video summaries are used as a fast way to reminisce about a past day. It emerged from our study that two minutes seems to be an appropriate duration.
2. **Show an image for 3 seconds.** Viewers need time to process a shown image. Images need to be recognized, understood, and reflected upon to activate the script and reconstruct memories through an inferential process. This allows an image to convey far more information than what it features.
3. **Images must be presented in their chronological order.** Showing images in the chronological order helps to understand the interdependence of events. Timestamps support this effect since viewers can relate the time to their script knowledge. By promoting the inferential process, the video summary would convey far more information than all single images of the video would do.
4. **Include images that feature persons, place, object and action cues.** These cues have shown to be the most effective to recall a past event. However, one type of cue is mostly not enough to be unambiguous. Hence, images that show more than one type of cue are required.
5. **Cues must be relevant and remarkable.** Shown cues must represent relevant information. People must be relevant to the event (i.e. the viewer must know that person or the person must affect the viewers event in question) while the location must be represented by a distinct environment or building. Distinct buildings mostly have a remarkable color layout or shape.

6. **From a set of similar images, select the ones that contain the most details.** The more details are shown, the more possibly useful details can be considered to recall a past event. Hence, from a set of similar images, select the one that shows the most details.

7. **Keep the context understandable.** The more changes occur between images, the more information must be provided so that the viewer can follow and understand the changes. Hence, the time-gap between shown images must adapt to the occurring changes (e.g. more images for moving events should be shown in comparison to non-moving events that are mostly represented by the same images).

8. **Avoid impaired images if possible.** Blurriness, camera lens occlusions through objects (e.g. hair, scarf) or an unfavorable viewing angles (e.g. relevant information is difficult to recognize) can confuse the viewer and require more time to recognize. It is also possible that viewers don't understand the given cue and lose their focus. However, impaired images must not be excluded completely due to the limitation of the lifelogging camera.

5.2. Design and Architecture

The limitations of lifelogging cameras resulted in many indistinguishable images that make it even difficult for humans to understand. Fulfilling aforementioned requirements presumes many experimentation and testing phases since many state-of-the-art solutions in the field of image processing are not designed to consider exceptional cases as such what we found in lifelogging images. Additionally, image processing operations such as a face detection on over 1,000 images are very time-consuming which hinders the experimental style of development. Hence, we had to add following additional requirement to our software development process:

9. **Component-Based Architecture:** The software must be designed with flexibility and a component-based architecture in mind to *(i)* enable testing and exchanging single components and *(ii)* cache results of time-consuming operations. This helps us to try out different approaches by exchanging the implementation of components with each other (and even combining them) and opens doors for further development.

In the following, we will first describe the system structure and then focus on implementation details of each of these components.

5.2.1. System Overview

Figure 5.1 shows the structure of the complete system. The user input consists of a set of NarrativeClip images (and associated metadata files), and a valid GPX File[1] that contains a recording of GPS Locations in the same period of time in which the NarrativeClip is operated. Since the image transfer from the NarrativeClip to a computer requires a proprietary software "NarrativeUploader"[2], we won't cover

[1] GPX 1.1 Schema Documentation: http://www.topografix.com/gpx/1/1/ (last accessed on October 10, 2015)

[2] NarrativeUploader Software Download: http://start.getnarrative.com/ (last accessed on October 10, 2015)

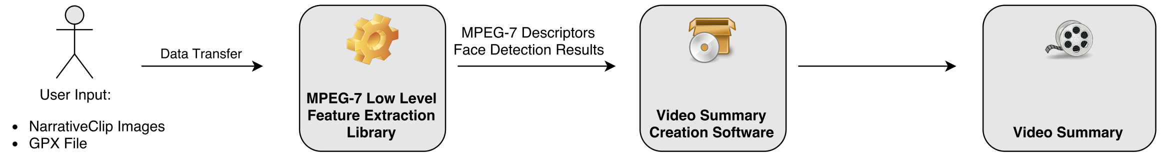


Figure 5.1.: An overview of the context of our video summary creation system. Captured images are first transferred to the MPEG-7 extractor. MPEG-7 descriptors are then passed to our software that creates the video summary.

this part in our software. The same applies to the GPS data, which can be collected with an arbitrary software or even hardware. In our work, we use the GPS Logger for Android[3].

Next, we run an own version of the MPEG-7 Low-Level Feature Extraction Library [BcGU09] written in C++. We extended it to extract metadata for all images in one folder and perform a face detection using OpenCV afterwards. The results are then written into a text file. This metadata includes four MPEG-7 descriptors that we will describe in the next section and the result of a face detection describing the position and size of faces (if detected).

The result of this extraction process is then passed to our software that we wrote in Java 8. This creates a video summary based on the metadata of the images and requirements described above. Separating the image feature extraction from our software allows us to test the software without having to wait for a time-consuming process (in our case it took about 4 hours for about 1,000 on an Intel i5 Laptop). The resulting video can then be watched with common media players such as VLC media player[4] or KMPlayer[5].

5.2.2. Software Architecture

Our software is composed of five main components; one model holding all data and four components that implement one step of the summary creation process as shown in Figure 5.2. The four steps strongly follow the process in which video summaries were created in our study: First, all collected data are gathered and prepared in a storage. Next, images are segmented into main activities of the day based on attributes and characteristics that we will present below. From these main activities, we select a few images that represent the activities appropriately (we refer to as *representatives*) and pass these images to our video creator to finally retrieve a video file.

The `model` package contains an `ImageStore` and a `LocationStore` to store pointers to images and all collected context data; and a `DissimilarityStore` that provide results of pairwise comparisons between images a and b, which we refer to as *Dissimilarity* $ds_{a,b}$. These dissimilarities contain e.g. comparison of image histograms or distances between the GPS Locations of two images specified in

[3] GPS Logger for Android: https://play.google.com/store/apps/details?id=com.mendhak.gpslogger (last accessed on October 10, 2015)
[4] VLC media player website: http://www.videolan.org/vlc/ (last accessed on October 10, 2015)
[5] KMPlayer website: http://www.kmplayer.com/ (last accessed on October 10, 2015)

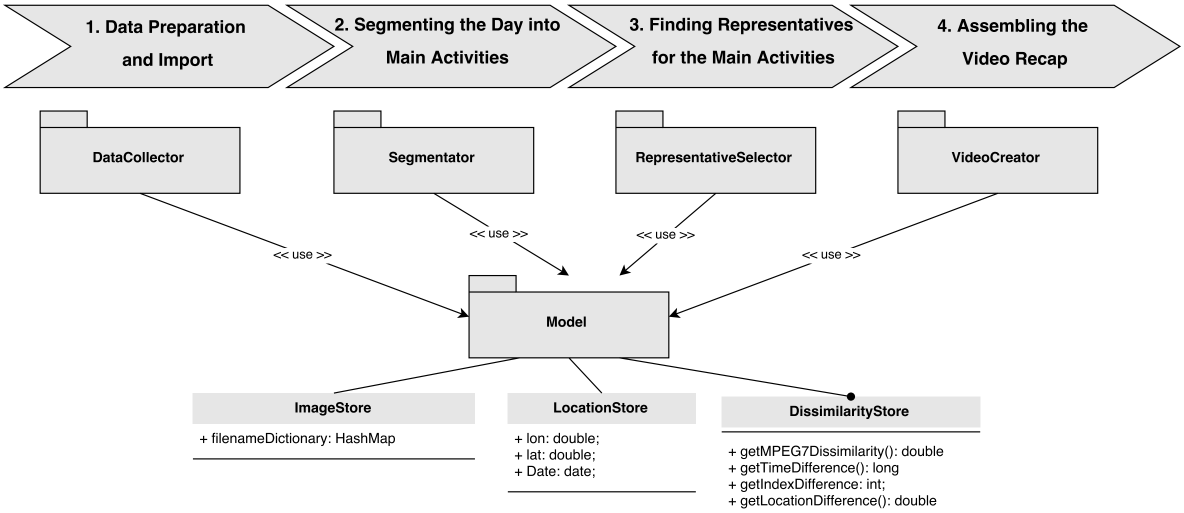

1. Data Preparation and Import
2. Segmenting the Day into Main Activities
3. Finding Representatives for the Main Activities
4. Assembling the Video Recap
DataCollector
Segmentator
RepresentativeSelector
VideoCreator
<< use >>
<< use >>
<< use >>
<< use >>
Model
ImageStore
+ filenameDictionary: HashMap
LocationStore
+ lon: double;
+ lat: double;
+ Date: date;
DissimilarityStore
+ getMPEG7Dissimilarity(): double
+ getTimeDifference(): long
+ getIndexDifference: int;
+ getLocationDifference(): double

Figure 5.2.: The architecture of our video summary creation software. At the top we can see the four steps of creating a video summary with the packages presented below.

kilometers. The motivation behind the `DissimilarityStore` is 1) to provide *min-max normalized* dissimilarity values for all pair of comparisons and 2) to reduce costs of repeated comparisons.

Instances of the `ImageStore`, `LocationStore` and the `DissimilarityStore` were filled with data by the `DataCollector`, which we will describe more detailed in the next section (5.4.1). Filled instances of those three classes serve as a database in all steps of the software and were hence passed to all components.

The remaining components `Segmentator`, `RepresentativeSelector`, and `VideoCreator` each contain the implementation of the segmentation, representative image selection and video assembly process. We will describe these components in the course of this chapter.

5.2.3. Frameworks

During the design phase, we looked at several frameworks and software libraries to cover the image processing functionality. We considered many alternatives and decided to use following three frameworks.

We used OpenCV, a library of algorithms aimed for support in computer vision, to cover the face detection functionality. Although the functionality of OpenCV is far too much for our needs, the face detection quality of simpler alternatives such as OpenIMAJ[6], [7] or lightweight ones such as JJIL[8]

[6] OpenIMAJ website: http://www.openimaj.org/ (last accessed on October 10, 2015)
[7] Marvin Project website: http://marvinproject.sourceforge.net/ (last accessed on October 10, 2015)
[8] JJIL website: https://code.google.com/p/jjil/ (last accessed on October 10, 2015)

couldn't convince us. Face detection algorithms of aforementioned libraries are based on a Haar feature-based cascade classifiers [VJ01], which in turn is based on pre-trained classifiers. In our tests with randomly selected images from participants, the pre-trained classifiers in OpenCV works the best.

To extract MPEG-7 histograms, we used the MPEG-7 Low-Level Feature Extraction Library [BcGU09], which is written in C++. We considered using LIRE[9] and the MPEG-7 tools Caliph&Emir [10], but decided against them due to being too overloaded or not providing all the type of histograms we wanted. In several related work, we noticed the reference of an MPEG-7 Extraction Toolbox named aceToolbox[DSLE07, BLBO$^+$06, BLD$^+$07], but couldn't find any source (except a dead URL) to download from.

Lastly, we used OGGSlideshow[11], which is a command line application to create slideshows. This allows a smooth integration into our four-step process and further provides the Ken-Burns transition effect[12] which is also used in Picasa as a transition effect for image slideshows.

5.3. Processing Features and Classification

To separate the image set into segments and select the most representative ones, we need a set of features on which we base the decision-making process of two said operations. In our implementation, we use the following four features that we describe below in more detail: *(i)* MPEG-7 descriptors, *(ii)* size of detected faces, *(iii)* NarrativeClip metadata, *(iv)* GPS data.

In this section, we describe the basics about these features before we describe their usage in the next section.

MPEG-7 Descriptor

We used the MPEG-7 Low-Level Feature Extraction Library [BcGU09] to extract four different MPEG-7 descriptors that we will describe in the following. MPEG-7 descriptors describe an image through histograms based on different features, such as the color layout or the edges found on an image. While three of the histograms we used (CLD, CSD, EDH) describe partitions of the images (here: blocks of the grid-separated image) with their bins, the CSD describe an aggregation of all available colors on the image. We extracted those histograms as vectors of numbers.

- **Color Layout Descriptor (CLD)**: Describes the spatial distribution of colors in an image that is partitioned into 8x8 blocks. In a nutshell, this can be imagined as an iconized representation of the image whereas each block is represented by an average of all its occurring colors.

[9] LIRE project website: http://www.lire-project.net/ (last accessed on October 10, 2015)

[10] MPEG-7 tools Caliph&Emir website: http://www.semanticmetadata.net/features/ (last accessed on October 10, 2015)

[11] OGGSlideshow website: http://www.streamnik.de/74.html (last accessed on October 10, 2015)

[12] Short explanation of the Ken-Burns transition effect on Wikipedia: https://en.wikipedia.org/wiki/Ken_Burns_effect (last accessed on October 10, 2015)

- **Color Structure Descriptor (CSD)**: The color structure descriptor describes an image by all its occurring colors. The extraction process goes through the whole image and counts the occurrence of specific colors in a histogram. This histogram is then condensed to 32 bins (despite other sizes such as 64, 128 or 256 are also possible, we chose the smallest one to be less strict on comparisons). [MVBE01]
- **Scalable Color Descriptor (SCD)**: The scalable color descriptor basically describes a color histogram which represents the color distribution in an HSV color space. The histogram values are normalized and nonlinearly mapped into a four-bit integer representation which gives a higher significance to smaller values [MSS02]. The SCD can be represented by histograms of 128, 64, 32 or 16 bins. We chose 16 bins to be less strict on comparisons.
- **Edge Histogram Descriptor (EDH)**: The edge histogram descriptor separates an image into 4x4 blocks and counts the edges (vertical, horizontal, 45-degree-horizontal, 135-degree and non-directional edges) in these blocks. Every bin of the histogram hence represent the number of edges found in one block.

Face Detection

We used OpenCV's built-in functionality to detect faces on the lifelogging images. OpenCVs face detection is based on object detection using Haar feature-based cascade classifiers [VJ01] and already ships with many pre-trained classifiers for faces. We used the face detection functionality with the following classifiers and parameters that Doherty et al. [DS08b] suggested in previous research: `haarcascade-frontalface-alt, scaling factor = 1.1, 3 neighbors, window size = 30 pixel`. We tried to adjust the parameters but couldn't find any improvement to the mentioned one.

NarrativeClip Meta Data

The metadata for NarrativeClip images are stored in JSON-Files in the `meta` folder on the same level as all images. The metadata of the NarrativeClip provide information about the context in the form of 11 attributes. However, 5 attributes are not documented at all so that we don't know what they indicate (`avg_win, awb_gain, aec_gain, aec_ecp, avg_readout`) and 2 of them are unimportant for us (battery level and firmware version). The remaining four are information about the magnitude, acceleration, trigger and lightmeter. While the first two seems to be interesting for our intentions, we found that resulting values are not reliable enough to use (i.a. too inaccurate and sometimes doesn't align with the image at all).

The trigger feature indicates whether an image was taken automatically (e.g. every 30 seconds) or manually triggered by the user. This is an interesting feature that we added to our set of features since a manually triggered image may indicate that the user wanted to take a photo of something interesting. The light meter feature describes the lighting condition of a taken picture which we also added to our set of features.

GPS Locations

We imported GPS Location from a GPX File[13], which stores GPS Locations and further metadata in a XML-based Layout. Since there is no out-of-the-box method to synchronize GPS points collected with the GPS Logger app to the lifelogging images, we had to implement our own synchronization mechanism using timestamps.

Every image got assigned a GPS point, that is time-wise the *closest* to the image while being logged *before* the image was taken. We decided to only select GPS points taken before the image to consider bad reception areas, such as rail tunnels. When the user enters a train and loses the GPS signal, all images taken after entry are assigned to the last GPS position (which is the starting train station). As soon as the user leaves the train and receives GPS signals again, the image after that will be assigned to the GPS Location of the ending train station. This allows us to detect a big jump location-wise and cover this case accordingly.

5.4. Implementation

After we described the features that gives the software an 'understanding' of the images, we will describe the implementation details of the four components in this chapter.

5.4.1. Step I: Data Preparation and Import

Input: *Paths to the images and context data.*
Output: *A model representing all collected data.*

The aim of this step is to import the images represented by MPEG-7 Histograms and associated context data into our model; namely the classes `ImageStore` and `LocationStore`. Further, dissimilarity values for all pairwise comparisons are calculated and results stored in the `DissimilarityStore`.

5.4.2. Step II: Segmenting the Day into Main Activities

Input: *`ImageStore` and `DissimilarityStore`.*
Output: *Image clusters representing the days main activities.*

The aim of this step is to cluster the images into segments that represent the day's main activities. This is done based on changes in MPEG-7 Histograms and the GPS Location. We tried out promising approaches from previous research first but weren't convinced by the results. This is why we developed our own algorithm.

[13] GPX 1.1 Schema Documentation: http://www.topografix.com/gpx/1/1/ (last accessed on October 10, 2015)

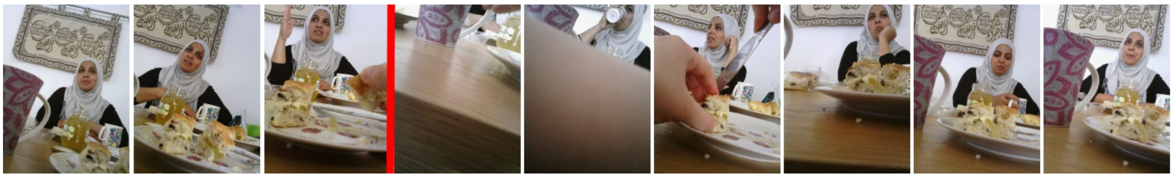


Figure 5.3.: Example of a wrongly identified event change due to changing the sitting position. The first and latter three images indicate that all nine images should belong to the same event while a peak scoring algorithm would detect an event change after the first three images. Found in an image set of P16.

Approaches from Previous Research

We tried out several algorithms from previous research, such as one promising approach from Doherty *et al.* [DS08a]. However, the approaches we found were developed and tested with images from the Microsoft SenseCam. Looking at example images of the Microsoft SenseCam, we realized that these were much more steady. We suspect that this may due to the SenseCam being worn with a necklace, which is more resistant towards body movements and loose clothes. Loose tops often result in unexpected changes of the vertical camera angle due to manual adjustment of users or by changing sitting positions. Further, the SenseCam has a fish-eye lense, which takes a bigger clipping and hence is more resilient against small movements.

The approach from Doherty *et al.* is based on an adaption of Hearst's TextTiling Algorithm [HP93]. This means that activity boundaries are triggered through big changes followed by smaller ones, which is detectable through peaks in a graph that plots the dissimilarity between the images n and $n-1$ [Figure 5.4]. Hence, this approach is prone to big sudden changes that last for a short period of time.

We have exactly this case when the NarrativeClip changes its angle due to users adjustments in a steady event (e.g. changing the sitting position or adjust the clothes). We show one example found in the image set of P16 in Figure 5.3. This shows 11 consecutive images in the chronological order, whereas the sitting position was changed for a short time after the third image. This triggered an event change in algorithms that are based on detecting these peaks. In the same paper, Doherty *et al.* suggested to consider blocks of images (an adaption of the TextTiling approach) instead of single ones and compare an average of them with each other. However, we still had an unpredictable behavior for such kind of cases.

Hence, we decided to develop out our own approach to consider these cases.

Clustering Algorithm

Our clustering algorithm is composed of two steps, which we will explain in the following. Opposing to Doherty *et al.*'s algorithm, we cluster images *without* considering the chronological order first. This removes the proneness to short interference scenes (e.g. image 4 to 6 in Figure 5.3) and allows us to focus on the image and context data exclusively. Thus, short and sudden changes, such as

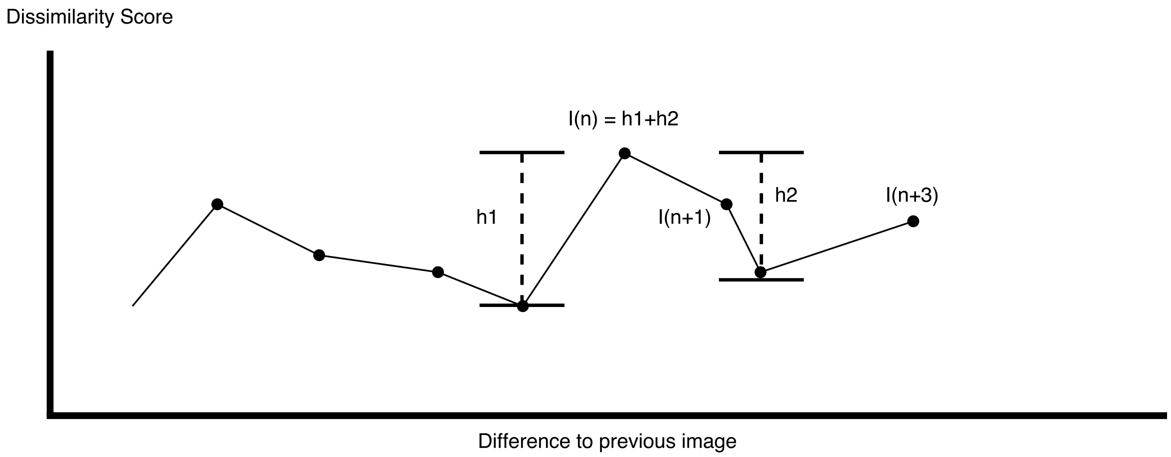


Figure 5.4.: Idea behind approaches that uses a peak in change to trigger event changes. The idea is to plot changes from image n to image n+1 and find the maximal turning point (here indicated through I(n)). Image as shown in [DS08a].

temporary changes of the camera angle or movements, won't trigger an event change anymore (in our example, the first three and last three would be assigned into one cluster). The chronological order will be recovered in a second step, in which we aim to fill eventually occurred gaps which supposedly represents the short interference scenes (the three images in the middle). By separating the clustering process into two steps, we gain control over short interference scenes for which we can then decide whether to trigger a new boundary for them or not.

Step 1. The first step is to cluster images based on their content (represented by MPEG-7 descriptors) and additional context data (GPS location). Although the k-Means clustering algorithm [SKK$^+$00] was shown to be one of the most popular approaches for clustering, we decided to implement our own algorithm to cover additional needs. We noticed that small changes over time, such as illumination, recording angle or slightly different positions of objects, often lead to a classification into different clusters. Further, assigning similar images from different parts of the day (e.g. sitting in the office in the morning vs. in the evening) to the same cluster opens more gaps than necessary. This would make the implementation of the second step more complex and error-prone.

In our algorithm, we iterate through all images and compare them to existing clusters. If a comparison does not exceed a certain threshold, the image is assigned to the cluster with the lowest dissimilarity. In case the threshold is exceeded, a new cluster is created for said image. A comparison of an image i to a cluster c is an aggregation of single comparisons of image i to every image j in cluster c, whereas newly added images are stronger weighted than the ones farther in the past. This is to address the problem of small changes over time. A pseudo-code demonstrating the comparison by calculating the dissimilarity is shown in Figure 5.1.

Further, an image i is automatically rejected by cluster c if i was taken later than one hour or more after the latest image in cluster c. This is to avoid assigning similar images from different parts of the day into the same cluster. We experimented with different time gaps and found empirically that 60 minutes seems to be the most appropriate value for this.

In the following, we present the implementation of the first step:

Algorithm 5.1 Algorithm to calculate the dissimilarity score between one image and a cluster.

```
double calcDissimilarity(image, cluster, startWeight, endWeight) {
  increment = (startWeight - startWeight) / imageStore.size();

  // calculate the average dissimilarity score between given image
  // and images of given clusters.
  weightSum = 0.0;
  dissimilarityScore = 0.0;
  for (int i = 0; i < cluster.size(); i++) {
    weight = startWeight + (i * increment);
    weightSum += weight;
    dissimilarityScore += weight * calcMPEG7Dissimilarity(image, cluster[i]);
  }

  // In case the given image is taken one hour or more after the latest image in
  // the cluster, we exclude it and put it into another cluster.
  // By returning MAX_VALUE, we guarantee that it will always be bigger than
  // any threshold.
  if (image.date - cluster[last].date > 1h) {
    return MAX_VALUE;
  } else {
    return dissimilarityScore / weightSum;
  }
}
```

Step 2. In this step, we aim to recover the chronological order by fixing occurred gaps. With gaps, we mean the set of missing images that are located, time-wise, between two images and are currently not in the same cluster as the two images. Gaps can be closed by either moving the missing images into the cluster or by splitting the cluster at the gap into two new clusters. In the following, we will first present five operations used to fill the gaps and then explain how we combined them to recover the chronological order.

We described the operations by a notation in which the letters A, B, and X represent different parts of a cluster. The letter A represents all images that time-wise occur *before* the gap and B all images that time-wise occur *after* the gap. The letter `X` represents all images that would belong into the gap time-wise. The notation of the operations describe how the cluster will look like after performing the respective operation. Brackets represent a new cluster (i.e. the old cluster is split up).

In the following we present the five operations and explain them in detail:

- `[A][X][B]`: This operation assigns A, B and X into three different clusters. This means that the original cluster is split into two parts (A and B), while all images that belong to the gap (X) are merged together into one cluster.
- `[A][XB]`: This operations assigns A into one cluster, and merges X and B into another cluster.
- `[AX][B]`: This operations merges A and X and assigns them into one cluster while B is assigned into another cluster.

- `[AXB]`: This operations merges all parts into one single cluster.
- `[A][Xm][B]`: This operations splits A and B to different clusters while splitting the images of the gap Xm into three activities. This operation will be described below in more detail.

To use these operations to recover the chronological order, we iterate through all clusters to find gaps and perform a decision-making process to decide which operation we use to fill the gap. The decision-making process is presented by an activity diagram in Figure 5.5.

Our decision-making process looks at the number of missing images (`|X|`) first. If the amount is less than 10, we perform operation `[AXB]`. This covers the case, in which a short interference of approximately 5 minutes resulted in assigning those images into another cluster (e.g. users blocked the lens with their arm or turned around). In case the gap requires more than 10 images to fill (means longer than 5 minutes), we make a decision based on from how many different clusters we have to gather those image from. Basically, this tells us whether the missing images were recognized as one single activity or as many different ones (respectively as one where the user was moving). In case the number of different clusters is exactly one, we can assume that this is most likely one particular activity (e.g. the user went to another room to talk with colleagues). In this case, operation `[A][X][B]` is chosen since we would merge two different activities into one cluster otherwise.

If missing images are from more than one different cluster, things get a bit more complicated. In general, we can assume that users are constantly moving when this is the case (e.g. going to another location). However, we have to check whether they stopped in between to perform a short, but maybe important, activity. This is done by ordering the missing images chronologically and checking whether there are more than 10 consequent images from the same cluster. If this is not the case, then we can now say with a high likelihood that the user was moving constantly and hence put those images into one cluster with `[A][X][B]`. If this is not the case, we now have the following constellation for X: `[moving 1][short activity][moving 2]` (Note: we except the case of two or more short activities, since we are working with gaps of at most one hour in which it is very unlikely to have more than one activities of at least 30 minutes). Here, we will first perform the operation `[A][Xm][B]` on the missing images X to separate the moving activities from the short, steady activity Xm. Afterwards, we perform again `[A][X][B]` on the cluster, so that the final result is: `[A][moving 1][short activity][moving 2][B]`.

Step 3. Our five operations presented above are moving images through different clusters. This may lead to a few clusters that contain a very small amount of images (less than 10) that may not be worthy to be considered as one single activity. In many cases, these are the beginning/ending of adjacent activities. This step is about fixing this limitation.

Basically, we iterate chronologically through the clusters and look for those very small clusters. We compare them with the last images of each of their adjacent clusters and calculate the dissimilarity score. In case the dissimilarity score stays under a certain threshold, we merge the small cluster with the neighbor with the lowest dissimilarity.

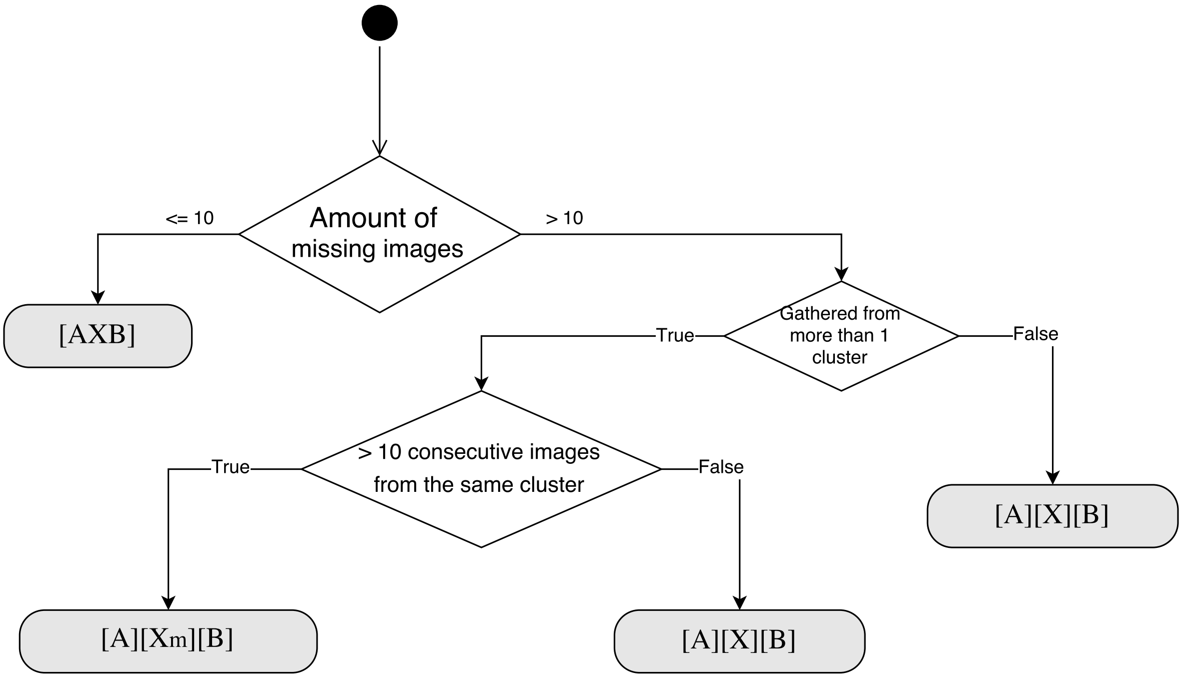


Figure 5.5.: Activity diagram of our decision making process to fill gaps with missing images.

5.4.3. Step III: Selecting Representatives for the Segments

Input*: A segment of images that represents a days main activity.*
Output*: One or more images representing the given segment.*

The aim of this step is to select representative images of a given segment; more exactly, landmark images that represent the given activity. Representative images are detected based on a novel metric that we will present in the following. Moreover, we define two additional constraints for the selection of representatives: *(i)* we aim to avoid picking nearly identical images as they are less likely to supply more information than different images would and *(ii)* in case more than one image is about to be selected, an appropriate time-gap between them should exist to represent different parts of the event.

In the following, we will first describe our metric which we refer to as relevance score, and will then explain how we used this metric to achieve the aim mentioned above.

Relevance Score Metric

While the implementation of some requirements seem to be clear and feasible (e.g. face detection, removing blurry and empty[14] images), there are also requirements which needs research work to be

[14] With empty images, we mean either completely black images or erroneous violet images with no content at all.

implementable. Examples are the remarkability of buildings, the amount of information found on an image and the recognizability of content due to lightning.

We tried out different features including results from image processing techniques, context data such as the acceleration and NarrativeClip metadata such as lightmeter and the trigger type (*"double-tap feature"* or time-triggered). For that, we rendered multiple sets of lifelogging images captured by our participants and ordered them by the feature score (e.g. from high lightmeter values to low lightmeter values). This allowed us to understand what these features do and how we can use them to approximate the relevance.

At the end, we have four features of which we use their aggregation as our measure for relevance. The features' results are normalized so that all features have an appropriate and controlled level of influence on the final representative score. These features are the amount of faces f, a graphical representative score based on MPEG-7 Histograms GRS, whether an image was manually triggered or automatically taken by the timer m, and a score lm representing the light meter feature of the NarrativeClip metadata. The final relevance score metric is calculated as follows:

$RS = f + GRS + m + lm$

Each part of this score RS will be described in the following.

Face Detection (f): Our first feature results from the face detection. For every face (up to 3) found in the image, the face detection score increases by 1. However, found faces have to bigger than 15% of the image size to be considered. The reason for this is to avoid considering people that are somewhere in the background and hence not interesting.

Lightmeter Score (lm): The lightmeter is a feature we found in the NarrativeClip metadata that indicates the illumination the image was taken with. However, there is no official documentation on any of the values found in the NarrativeClip metadata, so we don't know what the value exactly describes. The only confirmation that we have, are the empirical test we did before considering this feature.

Our analysis reveals that the lightmeter has a broad range of about 500–3,000 for images taken inside and about 30,000–150,000 for images taken under the influence of sunlight. While larger values don't affect the image negatively in terms of recognition, we noticed that values below 200 are correlating with dark and blurry images that are rather difficult to recognize. However, due to the limitation of capture devices, we didn't want to classify images as useless since there is still a chance that they contain a relevant memory cue that lacks alternatives.

Hence, we decided to score the lightmeter values as follows: 0 for 0..200, 0.5 for 200-500, 1 for > 500.

Graphical Representative Score (GRS): In chapter 5.3 we learned that bins in an MPEG-7 histogram represent particular areas on an image; e.g. in case of a Color Layout Descriptor, 64 bins describe an image partitioned into 8x8 blocks. A big variation between those bins indicates that the image shows different colors on different areas which most likely represent some information. In contrast, no variation would mean that the image just shows one single color. Transferred to our use case, a big variation between those bins implies a variety of information such as objects, people or buildings in different colors, whereas small variations correlate with monotonous images that usually show plain surfaces.

We continued this thought and empirically found three MPEG-7 histograms that enable us to approximate the amount of information and the remarkability of shown objects, which is embodied by notable colors and shapes. To be exact, we applied calculations on the color layout descriptor (CLD), color structure descriptor (CSD) and the edge histogram descriptor (EHD) and summed a normalized version of them up to a score that represents the graphical relevance GRS.

The final score is composed of the following parts, whereas h represents the respective histogram and $Var(X)$ the statistical variance of X.

Variance between bins of CLD: The variance between bins of the color layout descriptor indicates the colorfulness of an image. We calculated it as follows:

$$(5.1)\quad CLD_v = Var(h)$$

Variance of Differences between CSD Bins: The color structure descriptor represents an aggregation of a list of occurred colors. To use this information for our use case, we first calculated the differences d_i, j for all bins i to all other bins j, and then calculated the variance between all d_i, js. This describes the variation of colors on a given image. Hence, The formula looks as follows:

$$(5.2)\quad CSD_v = Var(\sum_i^{|h|} \frac{\sum_j^{|h|} abs(h(i) - h(j))}{|h|})$$

Mean of EHD bins: Lastly, we calculated the mean of all bins for the edge histogram. The edge histogram stores the amount of edges for every block, which means that the more edges we have, the more information is available in the image. The formula looks as follows:

$$(5.3)\quad EHD_m = \frac{\sum_i^{|h|} h(i)}{|h|}$$

Hence, the final score of this part looks as follows:

$$(5.4)\quad GRS = CLS_v + CSD_v + EHD_m.$$

Figure 5.6 shows an example of images in a descending order of our graphical relevance score. On the left, we can see images that are more likely to be selected due to the graphical relevance score than the ones on further on the right side.

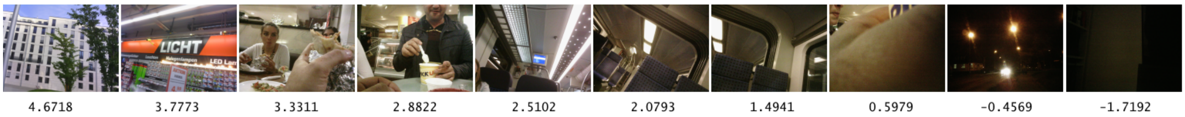


Figure 5.6.: A demonstration of our graphical relevance score. On the left we have images which have high graphical relevance scores and are hence more likely to be selected for inclusion in the video. Farther on the right we have rather bad images that are not likely to be included into the video.

Manual Triggering (mt): Manually taking pictures most likely means that users found something interesting that they want to keep. This is why the score for manual triggering is 1.0 when the image was taken manually, 0.0 when not. However, we noticed that many manually taken images were taken unintentionally – especially when the camera was adjusted or the wearer wore a loose top and was running. Unintentionally taken images have in common that they are blurry or show the sky/ceiling respectively the floor. Hence, we only considered an image as manually triggered when the graphical relevance score was over a certain threshold.

Selection of Representatives

Multiple nearly identical images are most likely not supplying more information than just one of them. Instead, they are taking slots away that could be filled with different and possibly more helpful images to trigger memories. Examples are images of the user sitting at the same place for multiple hours or even place the camera vertically on a surface. In this section, we describe how we used the relevance score that we described above.

The first step is to remove all nearly identical images while keeping just the best one of those. We do this by using a non-chronological clustering [Step 1 in the Clustering Component] with a much stronger threshold than before to assign nearly identical images into the same cluster. From each cluster, we keep the one with the highest relevance score.

Now that we only have diverse images representing one activity, the second step is to find the one with the highest relevance score. This is pretty easy to do since we just have to sort the images by their relevance score that we just presented above and pick the one on the top.

In case we have less than 10 diverse images, we are already done. If not, we have to find one more representative for every 10 diverse images. Since a big amount of diverse images most likely represent a longer lasting activity, it makes sense to select images that are more apart in terms of time (e.g. beginning and end of the activity). We attempted this by separating the set of diverse images into four quartiles and select images from quartiles that *(i)* are the farthest from each other and *(ii)* so that all quartiles are covered in case four or more representative has to be selected.

In the following we present our algorithm in pseudo-code:

```
public NCImage[] getRepresentativeImages(NCImage[] cluster) {
  Cluster representatives = {};

  // Step 1: Remove nearly identical images
  Cluster[] diverseImageClusters = segment(cluster, STRICT_CLUSTERING_THRESHOLD);
  Cluster diverseImages = new Cluster();
  for (Cluster c : diverseImageClusters) {
    sortImagesByRelevanceScore(c);
    diverseImages.add(c[0]);
  }

  // Step 2: Select the image with the highest representative score.
  sortImagesByRelevanceScore(diverseImages);
  representatives.add(diverseImages[0]);

  // Step 3: Select more images for every 10 more diverse images
  if (diverseImages.size() > 10) {
    for (int i = 0; i < diverseImages.size() / 10; i++) {
      NCImage nextRepresentative = getNextRep(diverseImages, representatives);
      representatives.add(nextRepresentative);
    }
  }

  return representatives;
}
```

Algorithm 5.2 Algorithm to select representative images from a segment.

5.4.4. Step IV: Assembling the Video

Input: *Representative images of all clusters.*
Output: *A video in the form of a slideshow presenting the representative images.*

We used the command line tool oggSlideshow to convert a set of images into a video in the OGV format[15]. The oggSlideshow command line tool expects a set of parameters[16], which consists of the video settings, the output file and a list of images to include. We run this tool with the following parameters:

```
oggSlideshow.exe -o output.ogv -s800x600 -l3 -tkb -d1024000 [<image.jpg>]
```

This creates a video file output.ogv (`-o` *`output.ogv`*) with a resolution of 800x600 pixels (`-s`*`800x600`*), an image display time of three seconds (`-l`*`3`*) and with Ken-Burns as a transition effect (`-t`*`kb`*). The `-d` switch sets the datarate in byte per seconds for the video encoder, which is a parameter we adopted from the websites examples.

[15] OVG is a video format especially used in the open source community. For more information, see its specification: https://tools.ietf.org/html/rfc5334#page-8

[16] oggSlideshow parameters and usage: http://www.streamnik.de/74.html

Additionally to the video file, a subtitles file (`*.srt`) is created that contains the timestamps. This file can be read by most media players such as VLC media player[17] or KMPlayer[18].

5.5. Summary and Discussion

In this chapter, we presented the requirements as an extract of the previous chapter and described our implementation to fulfill these requirements. Our software is composed of a model to hold captured data and four components that interact in a four-step pipeline with each other. This pipeline starts with a data import and preparation process implemented in the first component. Imported images and associated context data are then segmented into the main events of the day by the second component. A third component draw on these main events and selects the most important images from each segment which act as representatives for the event in question. These images are then passed into our fourth component that simply merges these images into a video and a subtitles file to display the capture time of the images.

The motivation for this component-based architecture comes from the desire of flexibility and extensive testability due to the sheer volume of techniques and solutions to implement the envisaged purpose of a component. The flexibility we achieved with this architecture further eases the future development as each component is an exchangeable part of the system.

Although we could have used observations of the video creation task in our study as ground truth data for a quantitative evaluation, we considered these to be very difficult to interpret in a meaningful way as there is not only one correct solution. Participants indecisiveness and not completely justifiable actions during the video creation task further confirm this. Hence, we decided to evaluate the software as a whole in the form of a qualitative evaluation presented in the next chapter.

[17] VLC media player website: http://www.videolan.org/vlc/
[18] KMPlayer website: http://www.kmplayer.com/

6. Evaluating the Software

This chapter is about the evaluation of our video summary creation system. The aim is to evaluate the effectiveness of created videos for reminiscence purposes and gather feedback on the current status. This feedback should be part of a solid base for the future work.

6.1. Methodology

This study is composed of two phases: We start with a briefing meeting and one day of capturing images with the NarrativeClip. This is followed up by another session including a semi-structured interview and questionnaires one week later.

6.1.1. Procedure

We designed this study with the aim to evaluate our software and gather feedback on the video summaries that were created by our software. After a short briefing on the usage and privacy implications of a 1st generation NarrativeClip prior to the study, participants recorded a full day and returned the camera on the next day. While returning the camera, participants were given the opportunity to delete images they didn't want to share with us. An appointment was then made for exactly one week later so that we have an eight-day period after which we could assume that participants had typically forgotten much of what occurred during the captured event [LC11]. The procedure until the evaluation meeting is exactly the same as in the first study to preserve the comparability between the recall performance measure.

The evaluation meeting is a semi-structured interview and covers two areas: *(i)* the recall performance evaluation and *(ii)* a feedback on the video summary. The recall performance evaluation remains the same as in the first study. We first asked participants to recall their day without any cues other than the date of the day they recorded one week ago. We then showed them the video summary and instructed them to use it as a memory aid to recall the recorded day. The above process was then repeated so that we can compare the results with each other to finally retrieve the improvement in recall.

In the feedback part of the study, we focused on three subject areas: *(i)* Usefulness and usability of a video summary, *(ii)* feedback on the video summary and *(iii)* suggestions for future work. The study was then closed up with a questionnaire about demographics.

6.1.2. Apparatus

Participants were issued with a 1st generation NarrativeClip to capture images of one full day. While capturing, participants additionally used a GPS Logger application[1] to log their location into a GPX File. The video summary was created with our video summary creation software.

6.1.3. Participants

We recruited 5 participants (4 male; average age = 26.4; SD=3.05) at the university in Stuttgart and from the circle of acquaintances of the author. Four of them are students in computers science while the remaining one was working as a banker. Participants were recruited by asking them personally and were rewarded with 15 EUR and a copy of all their collected lifelogging media at the end of the study. Prior to participation, 4 participants had reportedly never used lifelogging technologies while one person already worked with these technologies in his bachelors thesis.

The numeration of participants was continued from the last study. Hence, the first participant in this study is P17.

6.2. Results

In this section, we present the results of the recall performance evaluation, the feedback and suggestions for future work.

6.2.1. Measuring the Recall

An independent-samples T-test was conducted to compare the recall performance improvements triggered by the researcher-created video summaries in the first study and the summaries created by our software in this study. There was no significant difference between the recall performance scores for the researcher-created video summaries in the first study (M=126.53; SD=128.94; min=0; max=452.4) and the software-created video summaries in the second study (M=102.38; SD=94.52; min=17.9; max=238.1); t(19)=.385, p = .705. Levene's Test for equality of variances was found to not be violated for the present analysis, F(1,19)=.236, p=.633.

Participants reported that the video helped them to recall information that we coded into the following categories: *(i)* Locations, *(ii)* people, *(iii)* details and *(iv)* actions. Two participants reported that they forgot that they went to specific places, such as the supermarket (P18), the electronic market (P17) and even to a restaurant to meet friends (P17). Further, one participant admitted that she forgot show she met two acquaintances who ended up participating in her study (P18). Moreover, participants stated that they forgot details of their day, such as a laptop on the dinner table to show something to friends (P19) or that it was raining (P21). All participants admitted that they already forgot the duration and

[1] GPS Logger for Android: https://play.google.com/store/apps/details?id=com.mendhak.gpslogger (last accessed on October 10, 2015)

time of their activities and complimented the video on helping them to regain this information. P21 even recalled a wrong time first and corrected himself after seeing the video summary.

6.2.2. Video Summary as a Memory Aid

In general, participants rated the video summary as a helpful memory aid which is reflected by an average rating of 3.8 (SD=1.10) on a 5-point Likert scale (1="very bad"'; 5="very good"). We had 4 participants that would choose the video summary over a manual review of lifelogging images (equal to a 4 or 5 points on a 5-point Likert scale) while 1 participant had a slight tendency towards a manual review of captured images (2 on a 5-point Likert scale).

Participants complimented the length of the video which requires far less time than viewing all images manually. P17 further explained that "*although [viewing] images [manually] would supply [him] with more information, they also contain many boring ones*"[2]. Opinions of other participants (P19, P20) confirm this by commending the overview that the video gives in a short period of time ("*[The video] gave me a good overview. It described my whole day*"[3] - P19). P17 further likes the relationship between the images that make the day unique for him ("*It was really good. [..] I wouldn't confuse it with other days. [..] Through the relationship [between the images] many things become clear that wouldn't be the case when showing just single activities*"[4] - P17). This relationship also helped him to understand an image of a water tap that he complimented later to remind him of a special place ("*I liked [the image] with the strange water tap. With that, I knew that I was there*"[5]).

In terms of the selected images, 1 participant completely agree with how the video summary was created ("*[The video] was alright.*"[6] - P17) . 2 other participants agree with the content itself but would have liked to see more images on particular events ("*I think it wasn't bad. All main activities are included. Maybe I would have shown more of the dinner [..] since it was the most emotional activity on that day for me.*"[7] - P19). One remaining participant didn't like the images that were shown to her.

This participant (P18) stated that "*the video summary show many unimportant objects such as trees or part of the sky, means non-living objects*"[8]. Going through her images, she noticed that there were mostly useless images for recalling a past day, such as images of the sky, trees, the ceil or very blurry ones. Here she realized that the capture device seems to take very low-quality images and that she

[2] Translation from a german response: "Das Video ist ja kürzer als wenn ich durch alle Bilder scroll, praktisch als kurze Zusammenfassung. Bilder wären sehr sehr viele gewesen, die zwar mehr Details geben aber da waren auch sehr viele Langweilige mit dabei gewesen."

[3] This response was translated: "[Das Video] war kurz und gibt einen guten Überblick. Es hat meinen kompletten Tag beschrieben"

[4] This response was translated: "Das war echt gut. [..] ich würde es nicht verwechseln mit anderen Tagen. [..] Durch den Zusammenhang wird vieles klarer. Bei einzelnen Aktivitäten würde ich jetzt bspw. nicht wissen an welchen Tagen das war und was die Reihenfolge ist"

[5] This response was translated: "Das mit dem komischen Wasserhahn fand ich echt gut, da wusste ich dann, dass ich dort war"

[6] This response was translated: "Des Video war schon in Ordnung."

[7] This response was translated: "Ich fande [das Video] eigentlich gar nicht so schlecht. Die Hauptaktivitäten sind drin. Ich hätte vielleicht mehr vom Abendbrot reingebracht [..] weil das für mich an dem Tag das Emotionalste war."

[8] This response was translated: "Ich würde mir eher die Bilder anschauen, weil das Video viele unwichtige Dinge wie Bäume oder den Himmel gezeigt hat, also nicht lebendige Objekte."

didn't notice that the camera was facing up most of the time during the recording day. Consequently, she suggested some images as an alternative for the images of trees and skies which our algorithm didn't consider. Examples are blurry images of her notes, a clipping of her study apparatus that mostly shows cables or a photo that only shows a specific shirt without the person. This conforms with the image category we found in the first study that requires background knowledge to understand. Altogether, these reasons lead her to prefer images over our video summary.

The duration for every image was 3.0 seconds in our video summaries. Two participants completely agree with the duration ("*The duration was completely right/sufficient*"[9] - P17/P18), while three other participants wished the duration to be slightly longer due to them wanting to look at the time ("*It was a bit too fast for me because I wanted to see the times simultaneously*"[10] - P21).

All participants, however, criticized the images taken by the NarrativeClip. P18, for example, expected the images to show much more while P17 complains about the angle in which his images were taken.

6.2.3. Suggestions for Improvement

We asked our participants how the video could be improved to further support them on the episodic memory recall. Two suggestions for improvement were already mentioned when we asked them about the video summary as a memory aid. The suggestions were using a better camera and selecting more images for particular events (preferably events with people).

Outside of the image-only boundary, participants liked to have a short location identifier (such as "university", "restaurant" or street names). This would help them to recognize the location faster. P19 derived this suggestion from an image of an empty street where he admitted having troubles to recognize at the first glance. Similar to this, P17 suggested showing images of the buildings respectively their entries additionally to images of inside a building. He explained that he was shown images of him being in two different electronic markets. While he had troubles to distinguish them at a first glance, an image of the entry showing the name of the market would have avoided this problem.

As people are important memory cues, P19 suggested showing all participants of an activity either in one video frame as a collage of images or as thumbnails of persons additionally to the images. This would help to see one of the most important information at one glance. To further help to remember activities with people, P18 would like to have sound recordings of conversations additionally to the images. While images alone (especially bad images in her case) don't remind the viewer of what exactly happened, sound recordings would help here.

We asked participants whether they are interested in watching video summaries after every day. 2 participants expressed an interest, such as P17: "*At the end of the day, it would be fun to see what I did during the day. Just like letting the day pass in review*"[11]. For one participant, it would only

[9] This response was translated: "Die Dauer war genau richtig/komplett ausreichend"

[10] This response was translated: "Ich fand es etwas zu schnell, da ich gleichzeitig die Zeiten sehen wollte."

[11] This response was translated: "Ist schon lustig am Ende vom Tag zu sehen, was man so gemacht hat. [..] so Revue passieren lassen."

be interesting on special days while another participant wouldn't want to watch video summaries. However, all interest expressions were "‘conditional"’. While P17 rated the idea of video summaries as very interesting, the price for the camera would be too high for him and also too cumbersome when meeting people. Further, he is afraid that other people might have a problem with being recorded ("*For me it would be good, but not for others when they knew that they were recorded the whole time*"[12] - P17)

6.3. Summary and Discussion

Our analysis revealed that there is no significant difference between the recall performances after reviewing lifelogging images with researcher-created video summaries that were created in the five-week study and software-created video summaries that were created with our software. Although the assumption of homogeneity of variances was not violated, we had noteworthy less participants than in the first study.

The analysis of the interview shows that 4 out of 5 participants were satisfied with the video summaries and prefer them over a manual review. However, in comparison to the video summaries we created manually in the first study, we received more criticism as feedback from our participants. In many cases, this was due to the limitations of the capture device where we failed to find appropriate alternatives since bad quality images tend to be excluded in our current implementation. Although one participant (P18) liked the idea, she would rather use the entire lifelogging image set instead of the video summary. This is due to the majority of her images that feature ambiguous information (e.g. sky, trees, ceilings) for which she might need more information about the context to understand them. Other participants also criticized the clipping of the camera and suggested a fish-eye lens to maybe be more successful.

There were also suggestions on showing more images of people. Although relevant and valid faces are already strongly weighted in our selection algorithm, participants still found images that contain relevant people in their collected image set that wasn't considered by our algorithm. Our post analysis of the images that feature people they pointed out revealed that the face detection algorithm doesn't recognize them. Reasons are *(i)* the person on the image is not looking into the camera, *(ii)* only a part of the face is visible and *(iii)* they look into the camera in an unfavorable angle for the detection algorithm. Improvements or extensions (e.g. skin detection) may help us to yield a better result in the future.

Participants suggested innovative and reasonable ideas for improving the video, such as sound recordings, recognizing all present people or showing the entry of a building. Although we completely agree that these information will bring noticeable improvements to the videos, some of them require additional sensors in the camera or even additional devices to carry with. While an implementation may be feasible in future work, there is still a major challenge in finding an acceptable trade-off between more information and privacy limitations as well as practical limitations.

[12] This response was translated: "Also für mich wärs gut, aber für die anderen wärs nicht gut, wenn die wüssten, dass die die ganze Zeit aufgenommen werden"

Besides ideas for future work, there are also suggestions on improving our current system that do not require additional hardware. These include a better face detection, selecting more images for non-moving events and displaying the current location as a short address. Further, it may be worthwhile to consider a different lifelogging camera with a fish-eye lens.

7. Conclusion

This chapter summarizes the entire work and discusses the results that we found. Further, we present possible directions for future work.

7.1. Summary

Prior work has shown that lifelogging images support the episodic recall of not only the memory-impaired patients but also of the general population. However, the sheer volume of lifelogging images captured by lifelogging cameras exceeds the capability of users to review them on a daily basis. This deteriorates the perceived usability and discourages the general population from using lifelogging cameras as a memory aid. Hence, it would be desirable to select relevant images automatically and present them in a way that benefits the episodic memory recall.

In this thesis, we developed a software that creates video summaries of daily lifelogging image sets to serve exactly this purpose. We conducted a five-week study to elicit requirements from participants with the aim of informing the design of the system. Requirements were elicited through interviews and a task observation in which participants created a video summary using their own captured images. We found that images featuring a combination of people and location are the most effective cues. Additionally, our study shows that images should be presented in a chronological order and show distinct information to promote the inferential process [BEAA09, p. 180]. The inferential process reportedly enabled participants to reconstruct memories that are not directly featured in the video.

The system we developed is composed of four components that are strongly following the process in which video summaries were created in our study: (*i*) preparing the captured data, (*ii*) segmenting the day into main events, (*iii*) picking representative images for these events and finally (*iv*) create a slideshow video that present the representative images. We used MPEG-7, face detection and context data (i.e. GPS and NarrativeClip metadata) to approximate the similarity and relevance of the images.

The evaluation of video summaries revealed that there is no significant difference in the effect on the episodic memory in comparison to review methods that present the entire lifelogging image set (i.e. non-summarizing review methods). Moreover, participants prefer video summaries over said non-summarizing review methods due to a better usability which can play an important role in elevating this memory augmentation technology to a mainstream technology. Participants' feedback indicate that the video summaries created by our software are already valuable memory aids. However, there are still some limitations regarding the recognition of relevant images. This is due to the limitations of the capture device that either do not capture relevant information or capture them in a bad quality. Our current implementation tends to exclude images with a bad quality (e.g. blurry images, lens occlusions) that would otherwise convey useful information to support the episodic recall and delegates the reconstruction of these memories to the aforementioned inferential process.

7.2. Discussion

Although the majority of participants agreed that our video summaries are effective memory aids for recalling episodic memories, there are still limitations regarding the image selection algorithm.

On the one hand, we have the limitations concerning the technical side, such as a rather limited face detection due to the lack of pre-trained classifiers for lifelogging images or limitations of the lifelogging camera that leads to blurry images or missing information due to an unfavorable recording angle of the camera.

On the other hand, there are still image selection decisions of participants that we can not reconstruct and generalize despite extensive interviews, task observations and post analysis. We have this challenge especially for images that require background knowledge, such as low quality images, memories that are only valuable in conjunction with the specific context, or images that are included to promote the reconstruction of details (e.g. feelings in a specific moment). While we succeeded in creating video summaries to support the recall of main daily events, aforementioned limitations does not allow us to find appropriate cues to remind viewers of small but potentially important details. At the current state, we delegate this work to the inferential process of the viewer to reconstruct the missing memories that are not featured in the video. While many participants clearly confirm a successful inferential process one week after capturing the images, we doubt that this strategy will work after multiple years of time.

We believe, that the first step to take up this challenge is to get to know the participants on a more personal level. Insights about their daily routine, social circle, behavior, interests, life goals, and ambitions enables an analysis of the relationship between image selections and the participant himself. This opens the gate to many new features to analyze the selection, such as the novelty of events, relevance towards life goals or importance of encountered people. In this work, we gathered the life goals of participants using a questionnaire by Roberts *et al.* [RR00] with the aim to improve our understanding of their selection process. However, we couldn't find any relationship between the participants' image selection process and their life goals.

Our video summaries reportedly helped participants to recall memories that they already forgot one week after capturing. Nevertheless, we failed to find a statistically significant relationship between the review method and the impact on recall. However, general comments from participants indicated that they perceived the video summary as the more valuable memory aid. Results of questionnaires about the user experience [LHS08] and cognitive load [Gro88] are conforming with this. This may be the result of what felt like a personalized approach – unlike the non-summarizing review methods, the summary video appeared to participants to have been carefully planned to support their personal recall.

In total, we believe that we are already able to create effective video summaries to support the recall of a past day's main events automatically. However, there are still some limitations when it comes to recognize personal cues that requires background knowledge to translate into memories. The solution to this problem requires much more context information which is a topic that we suggest for future work.

7.3. Future Work

As we already mentioned above, our video summaries are focusing more on the main activities of the daily routine and rather ignore small, but potentially important details. While participants could reconstruct these memories using our given cues, it is not guaranteed that they are able to do this after a longer period of time. Hence, future work should investigate this topic and focus more on personalization and understanding the person for whom the video summary should be created.

We believe that more information about participants (be it through interviews or by data collection) would definitely help and improve the understandings for their decisions. What we already started with gathering life goals and ambitions of our participants should be continued on a broader basis. This can be either done by asking participants for more details about themselves, such as interests, behavioral aspects, relevant people, or important locations (what would be complicated to find participants for) or by conducting this study with closer friends and family members. Having enough information about them would allow us to assess the importance of e.g. people, locations or objects, relevance to life goals or to consider behavioral aspects during the analysis of their image selection and use them to select images for the video summary later on.

Implementation-wise, this can be further extended by considering more data to assess the relevance of certain images as a cue. While we only considered one image set at once in this work, considering image sets of weeks or even months would enable us to detect activities [DCC$^+$11], novelties [DS08b] or habits. Further, we could extend our current face detection into a face recognition that allows us to distinguish between different kind of people, such as friends, family members or new people. Other viable data sources are calendars, social networks (e.g. facebook, twitter) or communication logs on smartphones which would give us much more features to assess the relevance of certain images. Moreover, the importance of events may also be detected through biometric data (e.g. pulse, skin temperature, heart beat) which was already investigated by Sas *et al.* [SFR$^+$13] and shown to be useful. While we assume that this would be feasible technical-wise, we also have to consider the privacy and practical limitations for the users since this is more than what we can expect from them for a more effective memory aid. For the elevation of these technologies to the mainstream, this is definitely an important aspect to consider.

Selected images in our video creation process are presented in a slideshow presentation. Participants and interested parties suggested a combination of our video summary and timelapses, resulting in a timelapse that stops at relevant images for a certain duration. We agree with them and believe that this may fix the limitations of our algorithm to not include potentially important images. However we decided to stick to our initial approach and didn't implement the suggestion into our software since this approach may also bring some limitations of the timelapse with it. However, a combination of timelapse and our video summary is definitely worth to investigate in further work since there is a chance that we have the advantages of both approaches without the drawbacks.

A. Questionnaire

In the following we show the closing questionnaire for the five-week study. The survey was conducted using the Google Forms service [1]. The participant filled out the questionnaire after the last session on the computer of the researcher. The following is an export of the survey we created on Google Forms.

[1] Google Forms: `http://docs.google.com/forms` (last accessed October 10, 2015)

Questionnaire: Lifelog Summary

1. **Please select your gender**

   Mark only one oval.

   ○ male

   ○ female

   ○ prefer not to say

2. **What is your age?**

   ____

3. **What is the highest level of education you've completed?**

   Mark only one oval.

   ○ None

   ○ High-School

   ○ Bachelor's Degree

   ○ Master's Degree

   ○ Ph.D.

   ○ prefer not to say

   ○ Other: ______________________________

4. **What is your occupation?**

   Mark only one oval.

   ○ Self-employed

   ○ Employee

   ○ Student

   ○ Homemaker

   ○ prefer not to say

   ○ Other: ______________________________

5. **How many hours do you work per week?**
Mark only one oval.

○ 1 - 10 hours
○ 11 - 20 hours
○ 21 - 30 hours
○ 31 - 40 hours
○ 41 - 50 hours
○ more than 50 hours
○ Variable (I'm a student, freelancer, ...)

Experiences in lifelogging and video creation

6. **How often do you use lifelogging technologies?**
Mark only one oval.

○ I use it everyday.
○ I use it once or twice per week.
○ I use it once or twice per month.
○ I use it on rare occasions
○ I never used it before.

7. **Which of the following lifelogging devices/technologies do or did you use?**
Mark only one oval per row.

	I use it regularly	I use it sometimes	I used it once, but stopped using it	I never used it
Cameras (e.g. NarrativeClip or Autographer)	○	○	○	○
Fitness/Health tracker (e.g. FitBit, Sony SmartBand, Xiaomi Mi Band, ...)	○	○	○	○
Sleep tracking device (e.g. Zeo, WakeMate, FitBit, ...)	○	○	○	○
Location tracker (e.g. location tracking app such as PlaceMe, other tracker logging your location, ...)	○	○	○	○
Device to track your eating habits (e.g. HAPIfork)	○	○	○	○
Audio Recorder (e.g. Kapture Audio Recording Wristband)	○	○	○	○
Apps to log your usage of the mobile phone	○	○	○	○
Apps to log your usage of the computer	○	○	○	○
Diary	○	○	○	○
Scrapbook	○	○	○	○

8. **Do you use any other lifelogging devices? If yes, which?**

....

9. **How would you rate your experiences in video editing?**
Mark only one oval.

○ I create or edit videos professionally.

○ I create or edit videos regularly.

○ I create or edit videos sometimes.

○ I read something about video editing or tried it once.

○ I never created or edited videos.

Your experiences using the NarrativeClip

10. **How did people react to you wearing a NarrativeClip?**

1 = Nobody reacted like this; 5 Everybody reacted like this

Mark only one oval per row.

	1	2	3	4	5
They didn't notice the NarrativeClip.	○	○	○	○	○
They were comfortable that I used it.	○	○	○	○	○
They changed their behaviour after they know what the device does (e.g. became more conservative or shy and tried to avoid the camera).	○	○	○	○	○
They requested immediately that I deactivate it.	○	○	○	○	○

11. **What was your feeling on wearing a NarrativeClip over the day?**

Mark only one oval per row.

	Strongly disagree	Disagree	Agree	Strongly agree
I was afraid of recording things unintentionally, that I didn't want to record	○	○	○	○
I was afraid of losing/damaging the NarrativeClip while doing my activities	○	○	○	○
I was afraid of other peoples reaction on me recording them	○	○	○	○
I was afraid of covering the device unintentionally behind my jacket/scarf/...	○	○	○	○
It does not affect my day at all.	○	○	○	○
I paid less attention to certain things, because I felt like the NarrativeClip acts as a photographical memory for me.	○	○	○	○
I felt guilty, because I recorded people without their acknowledge.	○	○	○	○

12. **How did you view your NarrativeClip images in the last week?**

Please select "Other" if you used another approach or software to summarize your images and briefly describe how you did it.

Check all that apply.

□ Using a file manager to browse and an image viewer to view the images

□ Using the mobile app of NarrativeClip

□ Using the WebClient of NarrativeClip

□ I didn't view my NarrativeClip images at all.

□ Other: ....

13. **Has viewing NarrativeClip images helped you to improve the recall of your days?**
Mark only one oval.

	1	2	3	4	5	
No, not at all	○	○	○	○	○	Yes, immensely

14. **How much time did you invest into viewing your NarrativeClip images at the end of the day?**
Please enter the average amount of minutes per day.

....

15. **Did you enjoy viewing your daily images?**
Mark only one oval.

	1	2	3	4	5	
No	○	○	○	○	○	Yes

Content of a Daily Summary Video

16. **Which of the following information (found on your images) helped you to recall your day?**
1: Didn't help me at all; 5: Helped me immensely
Mark only one oval per row.

	1	2	3	4	5
Persons	○	○	○	○	○
Date (indicated by e.g. file name, lightning conditions, ...)	○	○	○	○	○
Weather	○	○	○	○	○
Place	○	○	○	○	○
Specific objects or buildings (e.g. a bag, a clock, a train station, ...)	○	○	○	○	○
Writings (e.g. text, numbers, or other readable information)	○	○	○	○	○
Whole environment of the image	○	○	○	○	○

17. **Are there other information on the images, that helped you to review your day? If yes, which?**

....

18. **Which of the following extra information would have improved your recall of your day?**

Those extra information would've been displayed together with the NarrativeClip images.
1 = Would've not helped me at all; 5 = Would've helped me immensely
Mark only one oval per row.

	1	2	3	4	5
GPS Data (represented as a map or the place/street name)	○	○	○	○	○
Time at which the image was taken	○	○	○	○	○
Events on your calendar or social network	○	○	○	○	○
Data about your movement (e.g. Accelerometer, Pedometer, ...)	○	○	○	○	○
Fitness/Health tracker (e.g. FitBit for measuring your sleep duration, burned calories, ...)	○	○	○	○	○
Important public news from that day	○	○	○	○	○
Indicating manually taken photos	○	○	○	○	○
Emotional state (e.g. sad, happy, angry, stressed, ...)	○	○	○	○	○
Smells	○	○	○	○	○
Audio Recordings (e.g. of conversations, of the sounds of the environment, ...)	○	○	○	○	○
Music you've listened to (retrieved from your media player)	○	○	○	○	○
Temperature	○	○	○	○	○

19. **Are there any other information that would have improved your recall of the day? If yes, which?**

....

20. **What is the maximum length of a daily summary video you would be willing to watch?**

Mark only one oval.

○ up to 30 seconds

○ up to 1 minute

○ up to 2 minutes

○ up to 3 minutes

○ up to 4 minutes

○ up to 5 minutes

○ more than 5 minutes

21. **Grouping images into clusters is a possible solution to summarize them and maintain a better overview. Which of the following grouping approaches would have helped you to review your day faster without losing too many information, that are relevant to you?**
1 = Would've not helped me at all; 5 = Would've helped me immensely
Mark only one oval per row.

	1	2	3	4	5
Grouping into periods (e.g. one partition per hour, one partition for morning/noon/evening, ...)	○	○	○	○	○
Grouping into events (recognition through calendar or social network events)	○	○	○	○	○
Grouping into places (using GPS)	○	○	○	○	○
Grouping into automatic triggered or manually triggered images	○	○	○	○	○
Grouping into people I met	○	○	○	○	○
Grouping into goals you followed with an action (e.g. being more sportive, progress in career, ...)	○	○	○	○	○
Grouping would not help me at all	○	○	○	○	○

22. **Are there any other methods of grouping to help you to maintain a better overview?**

....

Attitude towards lifelogging and its usage

In the following we will mention a daily summary video. Imagine a video, that is created by summarizing your NarrativeClip images and presenting them in the form of a slideshow.

23. **You want to achieve the following with the help of lifelogging:**
Mark only one oval per row.

	Strong disagree	Disagree	neutral	Agree	Strongly agree
Reminiscence (recall of memories)	○	○	○	○	○
Reflection (thinking about yourself and your actions)	○	○	○	○	○
Improving your time management	○	○	○	○	○
Create a good habit (e.g. eat healthy, going to sleep early, do more sports, ...)	○	○	○	○	○
Discard a bad habit (e.g. eat less chocolate, stop smoking, stop wasting time on something, ...)	○	○	○	○	○
Sharing with other people	○	○	○	○	○

24. **When would you watch a daily summary video about your day?**
Mark only one oval per row.

	Always	Most of the time	Rarely	Never
At the end of the day	○	○	○	○
At the beginning of the next day (e.g. directly after waking up, while sitting in the bus to work/uni, ...)	○	○	○	○
During the same week	○	○	○	○
During the same month	○	○	○	○
During the same year	○	○	○	○
Never	○	○	○	○

25. **With whom would you share your daily summary video?**
Check all that apply.

□ With my life partner (e.g. boyfriend/girlfriend or wife/husband).
□ With my family (e.g. brother/sister, mother/father, aunt/uncle, ...).
□ With my friends.
□ With my social network (e.g. facebook, Google+, ...).
□ With everyone.
□ I don't want to share it.
□ Other: ....

26. **Do you use social networks (e.g Facebook, Twitter, Google+, ...)?**
Mark only one oval.

○ Yes
○ I used them once, but I don't do it anymore
○ No

27. **How often do you post/tweet/broadcast something into social networks (e.g. facebook, twitter, ...)?**
Mark only one oval.

○ Regularly
○ Sometimes
○ Rarely
○ I use social networks, but don't broadcast anything
○ I don't use social networks

28. **Do you talk with your family/friends/partner about your day? (e.g. what you did, how you felt, how things went, ...)**

Mark only one oval.

○ Yes, I do that regularly

○ Yes. I do that sometimes

○ Yes. I do that, but only on rare ocassions

○ No. I don't do that

Bibliography

[ABB94] M. C. Anderson, R. A. Bjork, E. L. Bjork. Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1994. (Cited on page 11)

[ASC11] O. Aghazadeh, J. Sullivan, S. Carlsson. Novelty detection from an ego-centric perspective. In *Computer Vision and Pattern Recognition, 2011 IEEE Conference on*. 2011. (Cited on pages 14 und 34)

[BBK$^+$11] G. Browne, E. Berry, N. Kapur, S. Hodges, G. Smyth, P. Watson, K. Wood. SenseCam improves memory for recent events and quality of life in a patient with memory retrieval difficulties. *Memory*, 19(7), 2011. (Cited on pages 9 und 13)

[BcGU09] M. Baştan, H. Çam, U. Güdükbay, Özgür Ulusoy. BilVideo-7: An MPEG-7-Compatible Video Indexing and Retrieval System. *IEEE MultiMedia*, 17(3), 2009. (Cited on pages 47 und 49)

[BCM80] G. H. Bower, G. Clark-Meyers. Memory for scripts with organized vs. randomized presentations. *British Journal of Psychology*, 71(3), 1980. (Cited on page 43)

[BDSO08] M. Blighe, A. Doherty, A. F. Smeaton, N. E. O'Connor. Keyframe detection in visual lifelogs. In *Proceedings of the 1st international conference on Pervasive Technologies Related to Assistive Environments*. ACM, 2008. (Cited on page 14)

[BEAA09] A. Baddeley, M. Eysenck, M. Anderson, M. Anderson. *Memory*. Cognitive Psychologie. Psychology Press, 2009. (Cited on pages 11, 12, 21, 34 und 69)

[BGGU00] J. Boreczky, A. Girgensohn, G. Golovchinsky, S. Uchihashi. An Interactive Comic Book Presentation for Exploring Video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '00. ACM, New York, NY, USA, 2000. (Cited on page 15)

[BKW$^+$07] E. Berry, N. Kapur, L. Williams, S. Hodges, P. Watson, G. Smyth, J. Srinivasan, R. Smith, B. Wilson, K. Wood. The use of a wearable camera, SenseCam, as a pictorial diary to improve autobiographical memory in a patient with limbic encephalitis: A preliminary report. *Neuropsychological Rehabilitation*, 2007. (Cited on page 15)

[BLBO$^+$06] M. Blighe, H. Le Borgne, N. E. O'Connor, A. F. Smeaton, G. J. Jones. Exploiting context information to aid landmark detection in sensecam images. 2006. (Cited on page 49)

[BLD$^+$07] D. Byrne, B. Lavelle, A. Doherty, G. Jones, A. Smeaton. Using Bluetooth and GPS Metadata to Measure Event Similarity in SenseCam Images. In *IMAI'07 - 5th International Conference on Intelligent Multimedia and Ambient Intelligence.* 2007. (Cited on pages 14 und 49)

[BLJS08] D. Byrne, H. Lee, G. J. Jones, A. F. Smeaton. Guidelines for the presentation and visualisation of lifelog content. *iHCI 2008 - Irish Human Computer Interaction Conference 2008*, 2008. (Cited on pages 16, 34 und 42)

[BP$^+$06] M. Blum, A. S. Pentland, et al. Insense: Interest-based life logging. *IEEE MultiMedia*, (4), 2006. (Cited on page 14)

[Bus45] V. Bush. As We May Think. *The Atlantic Monthly*, 1945. (Cited on page 13)

[CF05] M. Cooper, J. Foote. Discriminative techniques for keyframe selection. In *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on.* IEEE, 2005. (Cited on page 14)

[ĆGC07] J. Ćalić, D. P. Gibson, N. W. Campbell. Efficient layout of comic-like video summaries. *Circuits and Systems for Video Technology, IEEE Transactions on*, 2007. (Cited on pages 9 und 15)

[CGL04] P. Chiu, A. Girgensohn, Q. Liu. Stained-glass visualization for highly condensed video summaries. In *Multimedia and Expo, 2004. ICME'04. 2004 IEEE International Conference on*, volume 3. IEEE, 2004. (Cited on page 15)

[CH02] M. Czerwinski, E. Horvitz. An investigation of memory for daily computing events. In *People and Computers XVI-Memorable Yet Invisible.* Springer, 2002. (Cited on page 13)

[Chi02] L. Chiariglione. Introduction to MPEG-7: Multimedia Content Description Interface. *Introduction to MPEG-7: Multimedia Content Description Interface*, 2002. (Cited on page 14)

[CJ10] Y. Chen, G. J. F. Jones. Augmenting Human Memory Using Personal Lifelogs. *Proceedings of the 1st Augmented Human International Conference*, 2010. (Cited on page 11)

[CJ12] Y. Chen, G. Jones. What do people want from their lifelogs? *Proceedings of the 6th Irish Human Computer Interaction Conference (iHCI2012)*, 2012. (Cited on page 27)

[CJG11] Y. Chen, G. J. Jones, D. Ganguly. Segmenting and summarizing general events in a long-term lifelog. 2011. (Cited on page 14)

[CLCC09] K.-Y. Cheng, S.-J. Luo, B.-Y. Chen, H.-H. Chu. SmartPlayer: User-centric Video Fast-forwarding. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09. ACM, New York, NY, USA, 2009. (Cited on page 15)

[DBS$^+$08] A. R. Doherty, D. Byrne, A. F. Smeaton, G. J. Jones, M. Hughes. Investigating keyframe selection methods in the novel domain of passively captured visual lifelogs. In *Proceedings of the 2008 international conference on Content-based image and video retrieval.* ACM, 2008. (Cited on page 14)

[DCC$^+$11] A. R. Doherty, N. Caprani, C. Ó. Conaire, V. Kalnikaite, C. Gurrin, A. F. Smeaton, N. E. O'Connor. Passively recognising human activities through lifelogging. *Computers in Human Behavior*, 2011. (Cited on pages 14 und 71)

[DS08a] A. R. Doherty, A. F. Smeaton. Automatically segmenting lifelog data into events. *WIAMIS 2008 - Proceedings of the 9th International Workshop on Image Analysis for Multimedia Interactive Services*, 2008. (Cited on pages 14, 52 und 53)

[DS08b] A. R. Doherty, A. F. Smeaton. Combining face detection and novelty to identify important events in a visual lifelog. *Proceedings - 8th IEEE International Conference on Computer and Information Technology Workshops, CIT Workshops 2008*, 2008. (Cited on pages 14, 50 und 71)

[DSLE07] A. R. Doherty, A. F. Smeaton, K. Lee, D. P. Ellis. Multimodal segmentation of lifelog data. In *Large Scale Semantic Access to Content (Text, Image, Video, and Sound)*. 2007. (Cited on pages 14 und 49)

[FBB11] J. R. Finley, W. F. Brewer, A. S. Benjamin. The effects of end-of-day picture review and a sensor-based picture capture procedure on autobiographical memory using SenseCam. *Memory*, 2011. (Cited on pages 9, 12 und 13)

[FFR11] A. Fathi, A. Farhadi, J. M. Rehg. Understanding egocentric activities. In *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011. (Cited on page 14)

[FO00] J. Farringdon, V. Oni. Visual augmented memory (VAM). In *iswc*. IEEE, 2000. (Cited on page 13)

[GAL05] J. Gemmell, A. Aris, R. Lueder. Telling stories with MyLifeBits. In *2005 IEEE International Conference on Multimedia and Expo*. IEEE, 2005. (Cited on page 13)

[GBL$^+$02] J. Gemmell, G. Bell, R. Lueder, S. Drucker, C. Wong. MyLifeBits: fulfilling the Memex vision. In *Proceedings of the tenth ACM international conference on Multimedia*. ACM, 2002. (Cited on page 13)

[GBL06] J. Gemmell, G. Bell, R. Lueder. MyLifeBits: a personal database for everything. *Communications of the ACM*, 2006. (Cited on page 13)

[Gir03] A. Girgensohn. A fast layout algorithm for visual video summaries. In *Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on*, volume 2. IEEE, 2003. (Cited on pages 9 und 15)

[GLB03] J. Gemmell, R. Lueder, G. Bell. The MyLifeBits lifetime store. In *Proceedings of the 2003 ACM SIGMM workshop on Experiential telepresence*. ACM, 2003. (Cited on page 13)

[Gro88] H. P. R. Group. NASA Task Load Index (TLX) v1.0. Paper and Pencil Package. *NASA Ames Research Center, Moffett Field CA*, 1988. (Cited on pages 18, 38 und 70)

[HHWH11] B. Höferlin, M. Höferlin, D. Weiskopf, G. Heidemann. Information-based adaptive fast-forward for visual surveillance. *Multimedia Tools and Applications*, 2011. (Cited on page 15)

[HP93] M. A. Hearst, C. Plaunt. Subtopic structuring for full-length document access. In *Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 1993. (Cited on page 52)

[HWB$^+$06] S. Hodges, L. Williams, E. Berry, S. Izadi, J. Srinivasan, A. Bulter, G. Smyth, N. Kapur, K. Wood. SenseCam: a retrospective memory aid. 2006. (Cited on page 21)

[KBH$^+$14] B. Kikhia, A. Boytsov, J. Hallberg, Z. ul Hussain Sani, H. Jonsson, K. Synnes. Structuring and Presenting Lifelogs Based on Location Data. In *Pervasive Computing Paradigms for Mental Health*, volume 100 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*. Springer International Publishing, 2014. (Cited on page 14)

[KHBS10] B. Kikhia, J. Hallberg, J. E. Bengtsson, S. Savenstedt. Building digital life stories for memory support. *International journal of Computers in Healthcare*, 2010. (Cited on pages 9 und 13)

[LC11] C. Loveday, M. A. Conway. Using SenseCam with an amnesic patient: Accessing inaccessible everyday memories. *Memory*, 2011. (Cited on pages 11, 18, 43 und 63)

[LD07] M. L. Lee, A. K. Dey. Providing Good Memory Cues for People with Episodic Memory Impairment. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '07. ACM, 2007. (Cited on pages 9, 13, 30 und 42)

[LD08] M. L. Lee, A. K. Dey. Lifelogging Memory Appliance for People with Episodic Memory Impairment. In *Proceedings of the 10th International Conference on Ubiquitous Computing*, UbiComp '08. ACM, New York, NY, USA, 2008. (Cited on pages 9, 10, 13 und 15)

[LF94] M. Lamming, M. Flynn. Forget-me-not: Intimate computing in support of human memory. In *Proc. FRIEND21, 1994 Int. Symp. on Next Generation Human Interface*. 1994. (Cited on page 13)

[LGG12] Y. J. Lee, J. Ghosh, K. Grauman. Discovering important people and objects for egocentric video summarization. In *2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, June 16-21, 2012*. 2012. (Cited on page 14)

[LHR$^+$09] S. E. Lindley, R. Harper, D. Randall, M. Glancy, N. Smyth. Fixed in Time and "Time in Motion": Mobility of Vision Through a SenseCam Lens. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '09. ACM, New York, NY, USA, 2009. (Cited on page 15)

[LHS08] B. Laugwitz, T. Held, M. Schrepp. *Construction and evaluation of a user experience questionnaire*. Springer, 2008. (Cited on pages 18, 38, 39 und 70)

[LKK11] M.-W. Lee, A. M. Khan, T.-S. Kim. A single tri-axial accelerometer-based real-time personal life log system capable of human activity recognition and exercise information generation. *Personal and Ubiquitous Computing*, 2011. (Cited on page 14)

[LL12] J. Ludewig, H. Lichter. *Software Engineering: Grundlagen, Menschen, Prozesse, Techniken*. dpunkt. verlag, 2012. (Cited on pages 17 und 24)

[LS15] M. Luchetti, A. R. Sutin. Measuring the phenomenology of autobiographical memory: A short form of the Memory Experiences Questionnaire. *Memory*, 2015. (Cited on pages 18 und 38)

[LSO+08] H. Lee, A. F. Smeaton, N. E. O'Connor, G. Jones, M. Blighe, D. Byrne, A. Doherty, C. Gurrin. Constructing a SenseCam visual diary as a media process. *Multimedia Systems*, 2008. (Cited on page 15)

[MGB12] M. Migueles, E. García-Bajos. The Power of Script Knowledge and Selective Retrieval in the Recall of Daily Activities. *The Journal of General Psychology*, 2012. (Cited on pages 12 und 42)

[MSS02] B. S. Manjunath, P. Salembier, T. Sikora. *Introduction to MPEG-7: multimedia content description interface*, volume 1. John Wiley & Sons, 2002. (Cited on page 50)

[MVBE01] D. S. Messing, P. Van Beek, J. H. Errico. The mpeg-7 colour structure descriptor: Image description using colour and local spatial information. In *Image Processing, 2001. Proceedings. 2001 International Conference on*, volume 1. IEEE, 2001. (Cited on page 50)

[PCS+12] D. Pavel, V. Callaghan, F. Sepulveda, M. Gardner, A. Dey. The story of our lives: From sensors to stories in self-monitoring systems. 2012. (Cited on page 13)

[PD+04] K. Peker, A. Divakaran, et al. Adaptive fast playback-based video skimming using a compressed-domain visual complexity measure. In *Multimedia and Expo, 2004. ICME'04. 2004 IEEE International Conference on*, volume 3. IEEE, 2004. (Cited on page 15)

[PEK+06] J. Pärkkä, M. Ermes, P. Korpipää, J. Mäntyjärvi, J. Peltola, I. Korhonen. Activity classification using realistic data from wearable sensors. *Information Technology in Biomedicine, IEEE Transactions on*, 2006. (Cited on page 14)

[PJH05] N. Petrovic, N. Jojic, T. Huang. Adaptive Video Fast Forward. *Multimedia Tools and Applications*, 2005. (Cited on page 15)

[RR00] B. W. Roberts, R. W. Robins. Broad dispositions, broad aspirations: The intersection of personality traits and major life goals. *Personality and Social Psychology Bulletin*, 2000. (Cited on pages 19, 41, 42 und 70)

[SB06] A. F. Smeaton, P. Browne. A usage study of retrieval modalities for video shot retrieval. *Information processing & management*, 2006. (Cited on page 14)

[Sch99] D. L. Schacter. The seven sins of memory: insights from psychology and cognitive neuroscience. *American psychologist*, 1999. (Cited on page 11)

[SFA+07] A. Sellen, A. Fogg, M. Aitken, S. Hodges, C. Rother, K. Wood. Do Life-Logging Technologies Support Memory for the Past? An Experimental Study Using SenseCam. *Chi '07*, 2007. (Cited on pages 9, 12, 13 und 21)

[SFR+13] C. Sas, T. Fratczak, M. Rees, H. Gellersen, V. Kalnikaite, A. Coman, K. Höök. AffectCam: arousal- augmented sensecam for richer recall of episodic memories. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*. ACM, 2013. (Cited on pages 13, 14, 21 und 71)

[SKK+00] M. Steinbach, G. Karypis, V. Kumar, et al. A comparison of document clustering techniques. In *KDD workshop on text mining*. Boston, 2000. (Cited on page 53)

[Squ92] L. R. Squire. Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *Cognitive Neuroscience, Journal of*, 1992. (Cited on pages 11 und 12)

[SW10] A. J. Sellen, S. Whittaker. Beyond total capture: a constructive critique of lifelogging. *Communications of the ACM*, 2010. (Cited on page 16)

[TT73a] E. Tulving, D. M. Thomson. Availability Versus Accessibility of Information in Memory for Words. *Journal Of Verbal Learning And Verbal Behavior 5*, 1973. (Cited on page 12)

[TT73b] E. Tulving, D. M. Thomson. Encoding specificity and retrieval processes in episodic memory. *Psychological review*, 1973. (Cited on page 12)

[Tul72] E. Tulving. Episodic and semantic memory. *Organization of Memory. New York: Academic*, 1972. (Cited on page 11)

[UFGB99] S. Uchihashi, J. Foote, A. Girgensohn, J. Boreczky. Video Manga: Generating Semantically Meaningful Video Summaries. In *Proceedings of the Seventh ACM International Conference on Multimedia (Part 1)*, MULTIMEDIA '99. ACM, New York, NY, USA, 1999. (Cited on pages 9 und 15)

[VJ01] P. Viola, M. Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1. IEEE, 2001. (Cited on pages 49 und 50)

[VSB06] S. Vemuri, C. Schmandt, W. Bender. iRemember: a personal, long-term memory prosthesis. In *Proceedings of the 3rd ACM workshop on Continuous archival and retrival of personal experences*. ACM, 2006. (Cited on page 13)

[WBH$^+$15] E. Woodberry, G. Browne, S. Hodges, P. Watson, N. Kapur, K. Woodberry. The use of a wearable camera improves autobiographical memory in patients with Alzheimer's disease. *Memory*, 2015. (Cited on pages 9 und 13)

[YKK$^+$09] K. Yasuda, K. Kuwabara, N. Kuwahara, S. Abe, N. Tetsutani. Effectiveness of personalised reminiscence photo videos for individuals with dementia. *Neuropsychological rehabilitation*, 2009. (Cited on pages 9 und 13)

All links were last followed on October 01, 2015.

Declaration

I hereby declare that the work presented in this thesis is entirely my own and that I did not use any other sources and references than the listed ones. I have marked all direct or indirect statements from other sources contained therein as quotations. Neither this work nor significant parts of it were part of another examination procedure. I have not published this work in whole or in part before. The electronic copy is consistent with all submitted copies.

place, date, signature