Impact of Video Summary Viewing on Episodic Memory Recall – Design Guidelines for Video Summarizations

Huy Viet Le[1], **Sarah Clinch**[2], **Corina Sas**[2], **Tilman Dingler**[1], **Niels Henze**[1], **Nigel Davies**[2]
[1]Institute for Visualization and Interactive Systems, University of Stuttgart
[2]School of Computing and Communications, Lancaster University
[1]{huy.le, tilman.dingler, niels.henze}@vis.uni-stuttgart.de, [2]{s.clinch, c.sas, n.a.davies}@lancaster.ac.uk

ABSTRACT
Reviewing lifelogging data has been proposed as a useful tool to support human memory. However, the sheer volume of data (particularly images) that can be captured by modern lifelogging systems makes the selection and presentation of material for review a challenging task. We present the results of a five-week user study involving 16 participants and over 69,000 images that explores both individual requirements for video summaries and the differences in cognitive load, user experience, memory experience, and recall experience between review using video summarisations and non-summary review techniques. Our results can be used to inform the design of future lifelogging data summarisation systems for memory augmentation.

Author Keywords
video summarization; lifelogging; episodic memory; recall; multimedia

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION
Tools and technology have long played a key role in supporting recall of events; photographs, drawings, diaries and other analogue records have allowed individuals to capture, preserve and reflect on memories. With the advent of digital tools, automatic capture, storage and presentation of memories is now also commonplace. Indeed, prior work has explored the use of specialized digital tools (lifeloggers) to enhance and augment memory (e.g. [26, 45]) with many envisaging a future in which technology automatically records and presents data about users' lives [13].

The rise of wearable and mobile computing, together with the commercialisation of lifelogging technologies and applications means that the creation of lifelogging records has become viable as a daily practice. Products such as Fitbit and Jawbone allow users to track activity and sleep patterns, whilst a wide range of mobile applications have emerged to support the tracking of daily activities such as locations visited, miles cycled, or coffees consumed. Visual lifelogging (i.e. the automated capture of visual images to record one's experiences) emerged through devices such as Microsoft's SenseCam. More recent devices (e.g. Autographer, Narrative Clip) add contextual data to captured images, for example location and motion patterns. Use of such visual lifelogging systems has a wide range of applications including self- and social-reflection [18, 25], health-behaviour monitoring [29, 40], autobiographical memory support [4, 26, 48], and general cognitive enhancement [46].

As we acquire the potential to gather increasingly large lifelogging datasets, managing them becomes a significant challenge. For visual capture devices such as the Narrative Clip, images may be captured at a frequency of one every 30 seconds for about 12 hours of the active day resulting in approximately 1,500 images per day. The sheer volume of this data exceeds the capabilities of users to review them all on a daily basis. Image processing may easily remove blurred, dark and indistinct images; however, providing appropriate review mechanisms is still challenging and is necessary for such devices to have the envisaged beneficial effects on human memory.

In this paper we consider a variety of options for the review of visual lifelogging data, with the specific goal of informing the design of automated systems capable of generating video summarisations of lifelogging images. We compare the experience and impact of viewing a manually-generated video summary with *non-summary* methods for reviewing visual lifelogging data (i.e. methods that present the entire lifelogging image set). Specifically our work compares video summaries with two non-summary approaches (timelapse video and manual browsing of thumbnail images) through a five-week user study involving 16 participants and over 69,000 images. Our study explores both individual requirements for video summaries and the differences in cognitive load, user experience, memory experience, and recall experience between review using video summarisations and non-summary review techniques.

CHI'16, May 07–12, 2016, San Jose, CA, USA

http://dx.doi.org/10.1145/2858036.2858413

The contributions of this work are as follows:

1. The results from a five-week explorative user study that identifies: *(i)* use cases for reviewing lifelogging images for memory augmentation, *(ii)* features of appropriate cues to be used in video summaries of lifelogging images, *(iii)* a structure for video summaries of lifelogging images, and *(iv)* the relative usability and effectiveness of video summaries for memory review when compared to non-summary approaches;
2. A description of a new metric, the *Recall Performance Measure (RPS_d)*, for evaluating improvements to recall for daily events; and
3. A set of requirements to help inform the design of video summaries of lifelogging images designed to support recall.

RELATED WORK

Lifelogging Data Capture

Early research into wearable computing proposed the idea of personal cameras as 'visual memory prosthetics' [39] and foresaw the era of visual lifelogging: real-life image capture through cameras embedded in glass-like prototypes being used to create a personal photo or video based memory prosthesis. One of the first feasible lifelogging cameras was Microsoft's SenseCam which comprised a sensor-enhanced image capture device worn around a user's neck to passively recorded images and context information at regular time intervals or based on changing sensor values (e.g. light, temperature). Research with SenseCam has demonstrated the utility of visual lifelogging in a range of application domains including self and social refection (for example in a work, education or family context) [18, 25], and monitoring of calorie consumption, exercise and other health behaviours [29, 40]. Focusing specifically on the use of lifelogging devices for human memory, case studies have shown that periodic review of SenseCam images can improve recall of personal events, providing autobiographical memory support for those with cognitive impairments such as amnesia and dementia [4, 26, 48]. Equally, the review of lifelogging data has been shown to act as a short-term general cognitive enhancement for those with no known memory impairments, improving performance in tasks involving episodic, working and autobiographical memory [46].

To investigate the efforts of digitizing an entire lifetime and storing related documents, Gemmell *et al.* [19] developed a lifelogging platform called *MyLifeBits*. In addition to timelapse footage from a SenseCam, they set out to capture a wide range of data including existing analogue content [20]. Chen and Jones [6] used an on-body timelapse camera with the purpose of supporting memory. Rather than using video footage, Chen and Jones argued for the use of photographs on the basis that watching video streams leads to information overload.

Managing Lifelogging Data

The volume of images captured by lifelogging devices quickly presents a challenge with regards to how to present the resulting datasets. Prior work has addressed this challenge either through the collection of additional context data during image capture to derive an image's significance later on, or through post-hoc image processing. Sas et al. [44], for example, presented a way of recording biometric data to later on filter images most relevant for users' recall performance. Gurrin *et al.* [24] record context data such as date, time, and GPS location for the organization of personal image collections to allow users to efficiently filter their photo archives. Post-hoc image processing and computer vision techniques can help filter out less-useful material by detecting over-exposed, dark and blurry images [9]. More sophisticated techniques include supervised learning to detect and recognize activities [16, 17, 43].

Related work has also looked into scene discovery [28], story-driven summarization [37], and key frame selection [14], as well as novelty detection in image sequences [1]. In these works, appearance and geometric cues based on alignment of the captured frame sequences of daily activities are combined with background deviation. Detection of human features can also be used to filter out relevant images from a stream of data: for example, faces, hands and object recognition based on foreground extraction [17, 34]. Such features can then be used again to understand activities [16]. Doherty *et al.* [15] built classifiers for visual lifelogs to infer personal lifetraits, such as people's characteristics and behavior. Hence, they were able to extract 22 distinct activities, such as meeting friends or having lunch.

Presentation of Lifelogging Images

Manual review of large quantities of lifelogging images is clearly infeasible – indeed current experiences with regular digital photography indicates that few people feel able to manage the volume of images generated.

A number of digital photograph curation systems have previously been proposed (e.g. [23]). However, for lifelogging review, simply managing the photos is not sufficient. Visually summarising a number of distinct images has previously been seen in timelapse photography in which many slowly-changing images are viewed in quick succession as a shorter video. Lindley *et al.* investigated this approach in a field trial with household members [35], whilst Berry *et al.* used timelapse videos as autobiographical memory aids for participants with memory impairments [2]. 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 various characteristics, such as information density [27], similarity measures [42], present motion [41] and manually defined semantic rules [7]. Similarly, image slideshows have been used as a video-based summary approach to help people with episodic memory impairment in the recollection of significant experiences [33].

An alternative approach for summarising images has used comic-like layouts [21, 47, 5]. Chiu et al. adapted these approaches and optimized them for mobile devices by using a voronoi-based layout [8]. Boreczky et al. further tried out these comic book presentations to navigate through videos

[3], while Lee *et al.* developed an interactive photo browser based on novelty values [31].

RESEARCH GOALS AND METHODOLOGY

We designed a five-week user study that aimed to both gather requirements for video summaries designed to support episodic recall, and to evaluate a manual implementation of these requirements. Our study specifically focused on the process of reviewing a single day's activities and looked to answer the following research questions:

RQ-1: what are the key features of daily lifelogging summary videos manually created by individuals to support review of events that occurred in their recent past (i.e. approximately one week ago)?

RQ-2: how do summary videos compare with non-summary approaches of reviewing lifelogging data for the review of activities that occurred in the recent 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?

Our study combined periods of lifelogging (during which participants were instructed to wear a camera device that automatically captured images) and a set of workshop sessions in which participants reviewed their images and participated in interviews. During the five-week study participants engaged in five lifelogging periods (each one day in length) and five workshop sessions (each approximately one hour in length), with the lifelogging days immediately preceding the workshop days. A one-week interval elapsed between each pair of study days [Figure 1].

Lifelogging capture days each followed the same procedure – participants were instructed to wear a 1st generation Narrative Clip[1] (a wearable lifelogging camera) that captured one image every 30 seconds when operational; the camera could also be manually triggered. The Narrative Clip device generates 5 megapixel images and incorporates 8 gigabytes of storage – allowing complete capture of at least one day. Participants were briefed prior to the study such that the privacy implications of the device (both for themselves and others) were clear and they were given advice on preventing capture of privacy-compromising situations (e.g. by covering the device or placing it in a pocket).

We ran four different types of workshop session. In each session participants first transferred images from their Narrative Clip to a researcher's computer. The sessions then continued as follows:

Session 1: Intro session During the introduction session participants were asked a series of questions about their prior experience with lifelogging, their ambitions and life goals, their most significant personal memories and their envisaged use of summary videos for memory review.

[1]http://getnarrative.com/

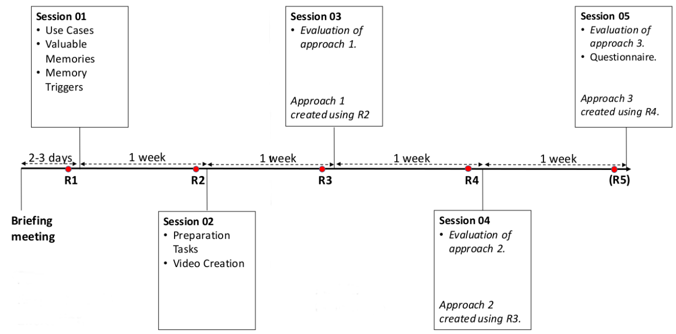


Figure 1: This diagram shows the process of the study that consists of five sessions. The boxes represent the session and the topics we focused on. The red points indicate the days R_i in which participants used the camera to record pictures. Recording always took place one day before a session.

Session 2: Manual video creation session Participants used Google's Picasa software[2] to complete a series of observed tasks using images from the lifelogging capture day that had occurred eight days previously[3]. Participants were first invited to review lifelogging images manually by browsing through the digital files using Picasa's grid view. Participants were then instructed to group their images into clusters – no specific direction for creating these clusters was given, participants were free to group images as they wished. Finally, participants were asked to create a video summary that would help them to remember the most important things from the captured day; participants were asked to 'think-aloud' as they completed this task. Videos were created as a Picasa slideshow by assigning pictures to an album and then defining the order and duration for each image. 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 and 4: Review session Participants were asked to review images from the lifelogging capture day that had occurred eight days previously. Images were reviewed using one of three methods: *manual browsing*, in which participants could browse through their digital image files using Picasa's grid view – a mouse enabled them to scroll through thumbnail images and to click specific images to view larger versions at a higher resolution – participants were not time-restricted in their review, but could work through images at their own pace (average time spent completing manual review was 153 seconds; min = 30, max = 445, s = 91.31); *timelapse*, in which all images captured by an individual were shown in order of capture in a timelapse

[2]http://picasa.google.com

[3]This recall interval was selected on the basis of prior research that indicates that recall failures increase significantly after a period of at least five days [36] – after eight days we could assume that participants had typically forgotten much of what occurred during the captured events and so we could test the effect of the review process.

video, each image was shown for the same duration and the total timelapse duration was 2 minutes; *video summary*, in which participants viewed a personalised video summary of their images that had been created by the researcher based on the outputs for that individual in the 'intro' and 'video creation' sessions. Researcher-created summaries were generated using the same approach that participants used in the video creation session: images were first clustered by activity and were then selected based on characteristics elicited from participants in their creation session. The mean duration of the videos summaries was 99 seconds (min = 54, max = 150, s = 24.14). Following the review activity, participants were issued with a series of questionnaires designed to evaluate the review method they had experienced.

Session 5: Closing session This session followed the same format as the review session (i.e. participants reviewed images using one method and then completed the evaluation questionnaires). Upon completion of the review tasks, participants were additionally issued with a closing questionnaire that revisited the use cases identified in the intro session.

The sessions ran in the order presented above with two review sessions (i.e. five sessions total). Over the course of the two review sessions and closing session every participant was exposed to all three review methods; the ordering of these methods was counter-balanced across participants.

Participants and Recruitment

We recruited a convenience sample of 16 participants (14 students; 2 staff) from the Lancaster University. Participants were recruited through email mailing lists and were reimbursed for their participation with £50 and a copy of their collected lifelogging images at the end of the study. Our participants were all aged between 18 and 39 years old (Mdn = 24.5; IQR = 6); 10 males and 6 females. Prior to participating in the study, 8 participants had reportedly never used lifelogging technologies, while 2 reported experiences in using lifelogging technologies on a daily basis.

Measuring Recall

In order to determine how well participants were able to recall days that they had summarised we developed an approach based on a method for memory probing described by Sas et al. [44]. In Sas et al.'s studies, participants were asked to recall the event, thoughts, emotions, place and time associated with a specific cue – answers were then scored with the value one if an answer was supplied and zero if it was not. Our approach builds on the work of Sas et al. but extends their method to evaluate the recall performance for a full day instead of just one activity and is specifically designed to compare recall before and after an intervention (in this case, image review).

Our approach, which we refer to as the *Recall Performance Measure* (RPS_d), is calculated as follows. First, we asked participants to recall their day without any cues other than the date of the day they recorded one week ago. We specifically asked participants to tell us the three most important events for them on that day and then probed for additional events if they were able to tell us at least three. We then asked participants for additional details for these three most important events; participants were asked to provide information about 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 the given answers to the details D_i with a score of either 0, 1 or 2 points. An answer was rated with 2 points when the answer is complete, 1 point when incomplete. Examples of 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 could not provide an answer. We subtracted 0.5 points from the score if participants hesitated while answering or inferred the answer from their daily routine or other activities (this was clarified by questioning participants).

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

$$ARS = \sum_{i}^{Details} \frac{\sum_{j}^{Events} Score_{i,j}}{|Events| * 0.14} \tag{1}$$

where

$$Score_{i,j} = D_{i,j} - P_{i,j}. \tag{2}$$

$P_{i,j}$ 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. Finally, to 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 the day d.

$$RPS_d = ARS_d * AR_d. \tag{3}$$

We then allowed participants to review their images (the intervention) and repeated the above process to get a second ARS value. To investigate the improvement in recall when calculating the second ARS value we focused on things they couldn't recall prior to the intervention and looked for corrections (of errors) in answers that were previously given. These errors were identified when participants admitted that they told us something wrong while recalling without the help of an intervention. We then calculated the difference between recall scores before and after the intervention (RPS_d).

The objectivity of the RPS_d measure was ensured through clear scoring criteria and was verified through a scoring of a random subset (25%) of participant responses by a second researcher; recall performance scores generated by the two researchers had a strong positive correlation, r = .778, p = .002.

RESULTS

Envisaged Use Cases for Lifelogging Summaries

When asked about the memories they most wanted to preserve, our participants were almost unanimous in their desire to remember positive experiences (*"Good times"* – P12, P15; *"Happy memories"* – P4, P7, P13; *"Precious moments"* – P8). Fifteen of our participants reported a desire to remember these positive experiences and many provided examples including travel experiences, special occasions, and time spent with important friends and family members. Ten participants identified a desire to preserve memories for social encounters more generally (*"Hanging out with friends"* – P8; *"life-experience related to socializing and meeting new people, making new friends"* – P3). Seven participants mentioned an interest in preserving memories for events that were somehow self-defining in nature: overcoming difficult situations, making mistakes, or goals that were achieved (*"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."* – P7; *"Things that make you feel as if you push the limits of life a bit"* – P6). Just three participants identified non-episodic memories as something they wished to preserve – these included knowledge from studies or conferences (P10, P1), the outcome of conversations (P11), and information about health behaviours such as eating, activity and caffeine consumption (P11).

Our participants expressed a range of motivations for preserving the selected memories. Many highlighted a relationship between their mood and recollecting a positive or negative experience – for some this was described as a process of actively drawing on memories to shape their mood (*"When times are sad, you can remember those happy moment and it cheers you up."* – P7) whilst others simply observed that recalling a positive memory was enjoyable and had the effect of improving their mood (*"It is nice to remember. . . "* – P16; *"They make me happy and feeling special."* – P5). Other common motivations included a desire to be able to draw on such memories to understand personal identity (*"Bad memories that you overcame makes you stronger. You know, if you've gone through a lot. . . it can make you a stronger person."* – P12), and an intention to reflect back in order to identify progress or points for improvement (e.g. *"[to] better yourself" – P6*). One participant also noted 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. Many participants also noted a link between their interest in preserving particular memories and their ambitions.

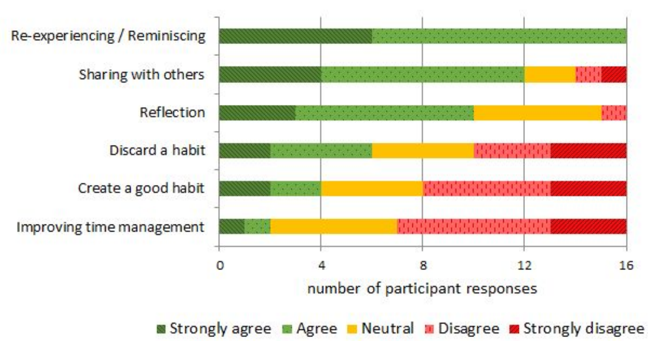


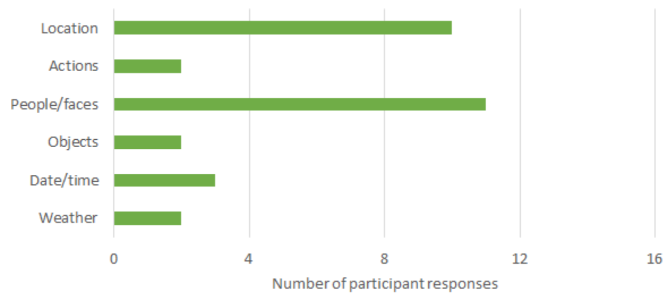
Figure 2: Likert-style responses to the question "You want to achieve the following by using a video summary of your day".

All participants reported an interest in using a summary video in order to re-experience or reminisce on their day. Twelve participants reported a desire to share such a video with friends and family (four additionally commented that they would be willing to share the video on social networks), and ten reported that they would use the video to reflect [Figure 2]. There was also interest in using videos for self-improvement, for example by altering personal habits or improving time management.

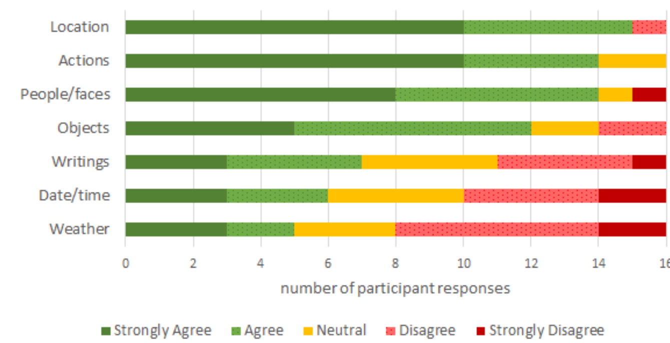
Cue Selection

We asked our participants to identify information that they considered would be beneficial for inclusion in an image if that image were to help them to re-experience an event. This was done at two stages in the study – in the opening session to identify promising cues, and in the closing session to detect changes in the elicited responses based on participants' experiences with lifelogging review. In the first session, we elicited 38 responses ($\bar{x}$ = 2.38 per participant, s = 0.86) which when coded equated to 15 unique answers. Coded answers fell into three categories: approximately half (7) described non-visual *contextual information and metadata*, for example date/time, weather, conversation, feelings and location; half (7) described *features of the image content itself* such as people, faces, landscape and objects; and one referred to the *framing of the image itself* ("the overall picture" – P15). The most popular responses are shown in Figure 3a and align with previous research from Lee *et al.* [32] who investigated memory triggers for people with episodic memory impairment. Most participants expected that images of others would be a useful cue – people (9 participants) and faces (2 participants); as would information that revealed the place in which an image was taken – location (10 participants) and landscape (2 participants). This combination of people and places was particularly common – for example, one participant noted that an image of a kitchen alone did not help him to recall a particular memory. Instead, he would also need the context in the form of people holding objects or laughing to be able to recall the specific memory. Indeed, as participants created their own video summaries we saw a common pattern of behaviour in which given a set of similar looking images, participants typically selected the one with the broadest combination of content types (e.g. a person *and* a recognizable location).

In the final session participants were asked to comment on the utility of each of the cues identified in session 1. This enabled us to explore changes in participants attitudes over the course of the study [Figure 3b] and we note that both people and locations remained popular as memory cues. However, by the end of the study both objects and actions were considered to be significantly more important as cues than they were ini-



(a) Cues identified in session 1.



(b) Cues rated in session 5.

Figure 3: Image features and metadata that could act as memory cues, as identified by participants in the first and fifth sessions. In session 1 responses took the form of coded free responses to an open question; in session 5 responses were Likert-style answers that indicated how useful they thought specific named cues would be.

tially; whereas only two responses for each were elicited for each in the first week, by the final week twelve participants agreed that objects would be a useful cue, and fourteen participants that actions would be useful.

If we look at the images selected by participants for inclusion in their video summaries, we see some deviation from the identified cues. Most notably, although people and locations seemed approximately equally useful to participants as potential cues, the actual occurrence of these differed significantly in participant videos. Whilst 45.0% of video images were selected to show a specific *location* exclusively (as seen in Figure 4a), only 23.2% show a person considered to be relevant to the captured event. This is most likely simply due to the lower occurrence of people captured in the lifelogging image sets – our observations showed that images containing relevant people were always selected when available.

11.8% of images from participant-created summary videos contain *objects*, such as food, laundry, presents or dishes. Participant comments indicated that these served two purposes: in some cases they were intended to remind the viewer of the objects themselves (e.g. lunch or gifts), whilst in others they were selected to represent an action, such as doing laundry or washing the dishes. A further 28.6% of video images contained *action* representatives [Figure 4b] in which actions are indirectly represented by persons, locations or objects. One observation from these images is that, whilst participants reported that they tried to select the most self-explanatory images, the limitations of the capture devices meant that participants often resorted to images that were only understandable with additional background knowledge. Figure 4c shows two such images – on the left a blurred image of a tree is used to remind one participant that they were hanging up their laundry, and on the right, the wooden texture reflects a period of guitar playing for another participant. These images are still effective cues since participants can recognize them and assign them to their original experience. This ability conforms to findings from Lee *et al.* [32]. In general participants selected better quality images where possible, avoiding those with blurriness, camera lens occlusions through hair, scarfs or other objects, or a bad recording angle.

Unsurprisingly, participants' cue selections tended to focus on the distinctive images of events that made the day different from other days. Examples are people that they normally wouldn't meet on that day or places like restaurants, meeting rooms or the homes of friends that they wouldn't visit everyday. Further, P11 stated that if she would watch the video everyday, she would only want to see the parts which are different from all the other days. This leads us to the conclusion, that participants need images that amplify what they already know about their daily routine. In this case, the context is needed as a link between the script and the potential missing piece of information from the video to recall a memory.

Video Duration and Structure

Our participants collected a total of 69,250 images in 80 image sets (16 participants with five lifelogging image sets each; $\bar{x}$ = 865.23 images per set, s = 418.9 images, min = 111, max = 1960). Most images were captured between 9am and 10pm, and across the participants we have coverage from everyday except Saturday.

Participants generally expressed a preference for moderately short video summaries. All of our participants would be willing to watch a video of no more than two minutes in length; six would be prepared to watch a video of up to three minutes in length, and three would be prepared to watch a five minute video. Indeed, when creating video summaries themselves participants tended towards even shorter videos; the mean duration for participant videos was 64.2 seconds (s = 35.9, min = 39, max = 188).

Participants filtered down from an average of $\bar{x}$ = 865.23 images per set to videos containing an average of $\bar{x}$ = 27.6 images (s = 21.03, min = 6, max = 81). Each image in participant-created videos was shown for an average duration of three seconds (the default value, no participant altered this) – many participants stated that substantial time was needed to recognise their images (P3, P12, P14, P15). Overall a dura-

(a) (b) (c)

Figure 4: Sample images from participants' video summaries showing (a) locations, (b) actions, and (c) images that require additional background knowledge to derive understanding,

tion of between 2 and 3 seconds was considered optimum by most participants.

All participant-created video summaries showed images in a chronological order (this was a selection participants made since no order was enforced on our participants) and participants often commented that this allowed them to make additional inferences about the events that occurred during the day. For example when watching a researcher-created video summary, one participant noted that since he wasn't shown revising in the library that day (his typical behaviour at the time) but was instead in another room to revise, he must must also have played football that day. This was not an activity he could recall engaging in, but he explained that he was revising in this room only because it was closer to the football pitch than the library is. Looking at the distribution of featured images, we find that participants did not select an equal amount of images from each portion of the day (as would occur in a simple timelapse). Instead we find that most (10) participants select more images for periods of the day 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% show only the route to the next location and hence are representing a movement exclusively. Although fewer images are used for static periods, three participants chose to deliberately select multiple images from static time periods in order to represent additional information not captured by any single image. 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 showed both her notes and the building. The majority of images either showed the building or the notes, but not both at once.

Comparing Video Summaries with Non-Summary Review

We compare participant experiences of using a researcher-generated video summary and the two non-summary review methods: timelapse and manual review. Participants were issued with questionnaires following experiences of each review method in order to establish the usability and utility of the methods for supporting recall. The results of these questionnaires are shown in Table 1.

Usability

Results from the NASA-TLX questionnaire [22] indicate that timelapse appears more cognitively demanding than manual review which in turn is considerably more demanding than video summary. This is supported by comments from participants regarding difficulties engaging with the timelapse and manual review processes. Many participants complained that the timelapse video was too overwhelming (P3, P11, P14), 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). P14 even stated that he would have been able to recall more if he had more time to think instead of being overloaded with images. Equally participants reported during the manual review process that they were "getting tired" (P3) and that the process was particularly time-consuming.

A one-way repeated ANOVA reveals a significant difference in NASA-TLX scores for the three review approaches, $F(2, 30) = 10.747$, $p<.001$. Bonferroni post hoc tests show a significant difference between the video recap and the timelapse video, $CI_{.95} = -37.591$ (lower) -8.534 (upper), $p = .002$ and between the video recap and manual approach, $CI_{.95} = -34.333$ (lower) -5.917 (upper), $p = .005$. No other comparisons were significant.

Analysis of Laugwitz et al.'s User Experience Questionnaire [30] scores show that, on all attributes, video recap yields a better experience than timelapse, which in turn provides a better experience than manual review. 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 recap and manual viewing approach. No other comparisons were significant.

Impact on Recall

Using Luttechi and Sutin's [38] 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 differ-

	Summary		Timelapse		Manual	
Improvement of RPS_d	126.53	(128.94)	151.60	(141.17)	77.19	(64.96)
Cognitive Load	21.00	(12.72)	44.06	(15.60)	41.13	(20.16)
User Experience						
Attractiveness	1.83	(0.86)	1.10	(1.16)	0.32	(1.08)
Perspicuity	2.00	(0.61)	1.38	(1.13)	1.25	(1.14)
Efficiency	1.42	(0.73)	0.95	(1.10)	0.61	(1.13)
Dependability	1.09	(0.60)	0.63	(0.83)	0.52	(0.78)
Stimulation	1.55	(0.96)	0.92	(1.04)	0.33	(1.19)
Novelty	0.86	(0.86)	0.36	(1.30)	-0.28	(1.33)
Memory Experience						
Vividness	2.25		2.13		1.63	
Coherence	2.09		1.94		1.97	
Accessibility	2.22		2.16		1.63	
Time Perspective	2.34		2.19		1.47	
Sensory Details	2.00		2.16		1.84	
Visual Perspective	2.31		2.06		1.63	
Emotional Intensity	2.00		2.19		1.81	
Distancing	2.06		2.00		1.94	
Preference	2.56	(0.63)	1.94	(0.77)	1.50	(0.73)

Table 1: Results of the Recall Performance Evaluation, NASA-TLX weighted rating, User Experience Questionnaire, Memory Experience Questionnaire (Mean Rank) and a score indicating which approach participants prefer. Participants ranked the approaches by their preference, which we rated with 3 points for the first choice, 2 points for second choice and 1 point for third choice. Values in brackets represent the standard deviation.

ences except for the time perspective, $\chi^2(2) = 8.415, p = .015$. No other comparisons were significant.

Overall our participants clearly felt that the video summary was a valuable memory aid. All participants reported that using the video they were able to recall the major events of the captured day and were therefore satisfied with the video recap as a memory aid. Participants commented that video summary was "very good" (P2, P8) and that "every major thing is captured in the video" (P6). P9 was even surprised by the result: "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.

Using our Recall Performance Measure we find that the largest difference (RPS_d) between recall performance before and after 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 approaches, $F(2,30) = 1.421$, $p = .257$.

Considering demographic issues (e.g. given psychology literature and general understanding that indicates memory changes as people age [12]), we find no significant result when using a two-way repeated measures ANCOVA to consider age as a factor, $F(2,28) = .611$, $p = .550$, nor do we find a significant relationship between hesitation and age ($r = .201$, $p = .455$). We note though that our participants were generally quite close in age (the IQR for participant ages was 6 years). We also fail to find a significant relationship between gender and recall performance ($r = .144$, $p = .329$). A two way repeated measure ANCOVA also confirms that session ordering has no impact on recall performance ($r = .172$, $p = .242$) suggesting that the study design was sufficient to prevent order effects from impacting our results.

ANALYSIS

Based on our experiences we can identify a set of requirements for video summaries of lifelogging events to support recall. These requirements can be used to inform the development of automated systems to generate such summaries and hence enable future memory systems to easily provide a mechanism by which users can review their experiences to enhance subsequent recall.

Our results also provide initial insights into the impact of summarizing lifelogging data into short, personalised, videos as opposed to more predictable timelapse approaches. Although there is limited evidence of tangible improvements in recall, video summaries were received more positively by participants suggesting that they are a valuable avenue for future research.

In the following sections we consider these two issues in more detail.

Requirements for Daily Video Summaries of Lifelogging Events

Our memories serve a range of purposes: helping us to regulate mood, maintain a sense of identity, build social relationships, and embark on range of self-improvement behaviours. The relative importance of different use cases is likely to vary over time, and thus the desire to preserve one memory over another is also subject to change. Our results indicate that although most participants were motivated by the desire to reminisce over positive experiences, this was not the only use case – all participants were able to list multiple use cases and motivations. From these observations we identify our first requirement: ***REQ 1*** *- video summarisations of lifelogging data should provide support for a wide range of use cases including reminiscence, self-reflection, promotion of self-identity and social interaction/sharing*.

There was clear consensus amongst our participants that images featuring depictions of people, places, objects and actions were most appropriate as memory cues. Location often occurred as a natural feature of images captured by the lifelogging devices. By contrast, images of people were much less common but were almost invariably selected for inclusion in video summaries. The importance of images featuring locations and people is reflected in the proportion of these cue types selected for inclusion in participants' summaries: 45.0% show a specific location while 23.2% show people (largely due to a lower incidence of capturing people than locations). Although not initially identified by participants as significant cues, objects and actions were a common feature of participants' video summaries indicating that these are often important for helping us restore the context of a particular event. Think-aloud comments from our participants showed that they would often recognise even heavily-degraded object representations (e.g. as a result of blurring, poor light, occlusion) and that these provided significant additional contextual information. We therefore suggest that ***REQ 2*** *- video summaries should promote images that feature person, place, object and action cues*.

In addition to the image content itself, many participants identified additional metadata that would add value to and improve the utility of a video summary. Although this metadata was not offered to participants for use in the creation of their summary videos, many of the identified data types are trivial additions to such videos. As examples, the date/time and location (GPS coordinates) for each image are typically readily accessible, web services could easily be used to add weather information (based on time and location), researchers have previously demonstrated the potential for connecting emotional state with lifelogging images [44], and snippets of audio have also commonly been associated with image files (e.g. AudioSnaps[4]). Our third requirement is therefore that ***REQ 3*** *- video summaries should incorporate a wide range of metadata to augment and annotate the visual content provided by the lifelogging images themselves*.

We saw a number of examples of participants reconstructing memories of events not directly featured in the videos. This was typically a result of either 1) a sense of chronology and interdependence between events (after this happened I would normally do that), or 2) additional knowledge about the significance of particular contextual cues (the pattern on that mug tells me I am working at home). This process of reconstruction allows the videos to convey far more information than they would do otherwise – each image need not form a complete memory in itself, but simply as to be sufficiently recognisable as to allow the viewer to access existing knowledge about their experiences. For this reason, we therefore identify a requirement that ***REQ 4*** *- video summaries should aim to provide scaffolding (e.g. chronology, context) to support the reconstruction of memories*.

With regards to video structure we have already noted the importance of chronology. The chronological nature of participants' summary videos might indicate that timelapse videos would be an acceptable substitute for personalised summary videos – if this were the case there are clear advantages since the process of generating a timelapse is straightforward and is identical for all users. Timelapse videos would also provide a clear representation of the proportion of time spent on specific activities since the number of images selected for any given activity would be direct reflection of the time spent engaged in that activity. However, analysis of participant videos indicated that although pictures were presented in chronological order some portions of the day were given a greater emphasis in summary videos than others. Long (static) events were rarely given due prominence in the videos (as they would for example in a timelapse) and activities with greater movement (e.g. walking between locations or playing sport) required more images to make the movements understandable. Furthermore, participants' opinions regarding repeated images for the same event were varied – some participants deliberately selected multiple images for an event in order to capture different aspects, whilst others commented that videos with multiple photographs of the same events "contained too many useless images". Knowing when and how to cluster multiple images for an event is a challenge to be addressed in future work, but existing literature about behaviour with other digital photography collections indicates that aspects such as time, location and appearance of images may be important [11]. Based on these observations, we therefore suggest that ***REQ 5*** *- video summaries should go beyond simply providing timelapse representations of events, and should only include multiple images for an event where these images provide additional details not represented well by a single image*.

Overall, regardless of the number and distribution of images selected, participants felt that a duration of 2-3 minutes was most appropriate for a video summary and most generated videos of 1-2 minutes containing approximately 20-40 images shown for three seconds each. This was comparable to the amount of time most participants spent manually reviewing images (1-4 minutes) suggesting that regardless of the review method most individuals will be unlikely to spend more than five minutes engaged in lifelogging review activities. Unsurprisingly we therefore suggest that ***REQ 6*** *- video summaries should be no more than 2-3 minutes in duration*.

Finally, our results indicate a clear improvement in user experience for video summaries over both timelapse and manual review. Comments from participants indicated that both video summaries and timelapse videos had the effect of providing participants with a more immersive media, and that manual review was not considered as positively due to the lack of a

[4]http://audiosnaps.com/

sense of immersion: "It doesn't feel like I'm the same person than in the video" – P10. The relationship between sense of presence and media has been well-explored, and richer media types are often credited with being more immersive and more effective at promoting a sense of presence. Our results suggest that this in turn may relate to the overall experience of engaging in memory review and we therefore propose that ***REQ 7*** *- video summaries should aim to immerse the viewer in their prior experiences through selection of sympathetic viewing angles (as suggested by [10]) and rich media use*.

Understanding the Impact of Video Summaries on Recall

Our participants' self-reports of their memory experiences (e.g. the vividness and accessibility of their memories) were more positive for video summaries than for other methods of review. However, a statistically significant difference for these memory experience reports was only seen around the timing of events – participants were more likely to report that their memory for the day and hour of occurrence of an event was clear following review by video summary than by other methods. General comments from participants also indicated that their perception was that the video summary was the more valuable memory aid. This may be the result of what felt like a personalised approach – unlike a timelapse, the summary video appeared to participants to have been carefully planned to support their personal recall. The video summary was also found to provide the better user experience, and this too may contribute to participants' perceptions that this review method was the more effective.

Despite participants' positive comments, we failed to find a statistically significant relationship between Recall Performance and review method. This may be a product of the small sample size or may indicate flaws in the video generation process, but clearly indicates that further work is needed to establish if the additional effort of generating a personalized summary video can yield benefits for recall.

CONCLUSION

In this paper we have presented the results of a user study designed to inform the production of video summaries of lifelogging data. Revisiting our research questions, we believe that our results do provide some insights into the key features of daily lifelogging summary videos created by individuals for review of events that occurred in their recent past (RQ-1), most notably the presence of people, places, objects and actions; a short duration; and a chronological structure. These features are important for understanding how to develop systems to develop such summary videos automatically, and we have presented a series of requirements for automatic summarisation systems based on these insights. Our results fail to provide evidence that summary videos are significantly more effective than non-summary approaches for improving recall (RQ-2A) but they do show that summary videos can provide an enhanced user experience (RQ-2B) – avoiding the high cognitive load associated with time-lapse and manual approaches to review.

Our future work is focused on using the requirements presented in this paper to inform the design and implementation of a system for automatically producing video summaries based on lifelogging data. As part of this work we are beginning to explore how we can separate out "special", noteworthy days from more ordinary days by recognizing changes in the lifelogging data captured. This is an important aspect of our overall drive to reduce the number of images that users must review in order to accrue the perceived benefits in memory augmentation.

We are also interested in exploring how we can utilise user stated objectives in terms of life goals and behaviour change targets to help inform the production of video summaries. In particular, we believe that using these objectives to help structure video summaries could lead to significantly higher quality video summaries.

ACKNOWLEDGMENTS

This research is partially funded through the Future and Emerging Technologies (FET) programme within the 7th Framework Programme for Research of the European Commission, under FET grant number: 612933 (RECALL).

REFERENCES

1. Omid Aghazadeh, Josephine Sullivan, and Stefan Carlsson. 2011. Novelty detection from an ego-centric perspective. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 3297–3304. DOI: http://dx.doi.org/10.1109/CVPR.2011.5995731

2. Emma Berry, Narinder Kapur, Lyndsay Williams, Steve Hodges, Peter Watson, Gavin Smyth, James Srinivasan, Reg Smith, Barbara Wilson, and Ken Wood. 2007. 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* 17, 4-5 (2007), 582–601. DOI: http://dx.doi.org/10.1080/09602010601029780

3. John Boreczky, Andreas Girgensohn, Gene Golovchinsky, and Shingo Uchihashi. 2000. 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, 185–192. DOI: http://dx.doi.org/10.1145/332040.332428

4. Georgina Browne, Emma Berry, Narinder Kapur, Steve Hodges, Gavin Smyth, Peter Watson, and Ken Wood. 2011. SenseCam improves memory for recent events and quality of life in a patient with memory retrieval difficulties. *Memory* 19, 7 (2011), 713–722. DOI: http://dx.doi.org/10.1080/09658211.2011.614622

5. Janko Ćalić, David P Gibson, and Neill W Campbell. 2007. Efficient layout of comic-like video summaries. *Circuits and Systems for Video Technology, IEEE Transactions on* 17, 7 (2007), 931–936. DOI: http://dx.doi.org/10.1109/TCSVT.2007.897466

6. Yi Chen and Gareth J. F. Jones. 2010. Augmenting Human Memory Using Personal Lifelogs. In

Proceedings of the 1st Augmented Human International Conference (AH '10). ACM, New York, NY, USA, Article 24, 9 pages. DOI: http://dx.doi.org/10.1145/1785455.1785479

7. Kai-Yin Cheng, Sheng-Jie Luo, Bing-Yu Chen, and Hao-Hua Chu. 2009. 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, 789–798. DOI: http://dx.doi.org/10.1145/1518701.1518823

8. Patrick Chiu, Andreas Girgensohn, and Qiong Liu. 2004. Stained-glass visualization for highly condensed video summaries. In *Multimedia and Expo, 2004. ICME'04. 2004 IEEE International Conference on*, Vol. 3. IEEE, 2059–2062. DOI: http://dx.doi.org/10.1109/ICME.2004.1394670

9. Sarah Clinch, Nigel Davies, Mateusz Mikusz, Paul Metzger, Marc Langheinrich, Albrecht Schmidt, and Geoff Ward. 2016. Collecting Shared Experiences through Lifelogging: Lessons Learned. *Pervasive Computing, IEEE* 15, 1 (Jan 2016).

10. Sarah Clinch, Paul Metzger, and Nigel Davies. 2014. Lifelogging for 'Observer' View Memories: An Infrastructure Approach. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (UbiComp '14 Adjunct)*. ACM, New York, NY, USA, 1397–1404. DOI: http://dx.doi.org/10.1145/2638728.2641721

11. Matthew L. Cooper. 2011. Clustering Geo-tagged Photo Collections Using Dynamic Programming. In *Proceedings of the 19th ACM International Conference on Multimedia (MM '11)*. ACM, New York, NY, USA, 1025–1028. DOI: http://dx.doi.org/10.1145/2072298.2071929

12. Fergus I Craik and Joan M McDowd. 1987. Age differences in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13, 3 (1987), 474.

13. N. Davies, A. Friday, S. Clinch, C. Sas, M. Langheinrich, G. Ward, and A. Schmidt. 2015. Security and Privacy Implications of Pervasive Memory Augmentation. *Pervasive Computing, IEEE* 14, 1 (Jan 2015), 44–53. DOI: http://dx.doi.org/10.1109/MPRV.2015.13

14. Aiden R. Doherty, Daragh Byrne, Alan F. Smeaton, Gareth J.F. Jones, and Mark Hughes. 2008. 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 (CIVR '08)*. ACM, New York, NY, USA, 259–268. DOI: http://dx.doi.org/10.1145/1386352.1386389

15. Aiden R. Doherty, Niamh Caprani, Ciarán í Conaire, Vaiva Kalnikaite, Cathal Gurrin, Alan F. Smeaton, and Noel E. O'Connor. 2011. Passively Recognising Human Activities Through Lifelogging. *Comput. Hum. Behav.* 27, 5 (Sept. 2011), 1948–1958. DOI: http://dx.doi.org/10.1016/j.chb.2011.05.002

16. Alireza Fathi, Ali Farhadi, and James M Rehg. 2011a. Understanding egocentric activities. In *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 407–414. DOI: http://dx.doi.org/10.1109/ICCV.2011.6126269

17. Alireza Fathi, Xiaofeng Ren, and James M Rehg. 2011b. Learning to recognize objects in egocentric activities. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference On*. IEEE, 3281–3288. DOI: http://dx.doi.org/10.1109/CVPR.2011.5995444

18. Rowanne Fleck and Geraldine Fitzpatrick. 2009. Teachers' and Tutors' Social Reflection Around SenseCam Images. *Int. J. Hum.-Comput. Stud.* 67, 12 (Dec. 2009), 1024–1036. DOI: http://dx.doi.org/10.1016/j.ijhcs.2009.09.004

19. Jim Gemmell, Gordon Bell, and Roger Lueder. 2006. MyLifeBits: A Personal Database for Everything. *Commun. ACM* 49, 1 (Jan. 2006), 88–95. DOI: http://dx.doi.org/10.1145/1107458.1107460

20. Jim Gemmell, Lyndsay Williams, Ken Wood, Roger Lueder, and Gordon Bell. 2004. Passive Capture and Ensuing Issues for a Personal Lifetime Store. In *Proceedings of the the 1st ACM Workshop on Continuous Archival and Retrieval of Personal Experiences (CARPE'04)*. ACM, New York, NY, USA, 48–55. DOI: http://dx.doi.org/10.1145/1026653.1026660

21. Andreas Girgensohn. 2003. A fast layout algorithm for visual video summaries. In *Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on*, Vol. 2. IEEE, II–77. DOI: http://dx.doi.org/10.1109/ICME.2003.1221557

22. Human Performance Research Group. 1988. NASA Task Load Index (TLX) v1.0. Paper and Pencil Package. *NASA Ames Research Center, Moffett Field CA* (1988).

23. Rebecca Gulotta, William Odom, Jodi Forlizzi, and Haakon Faste. 2013. Digital Artifacts As Legacy: Exploring the Lifespan and Value of Digital Data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 1813–1822. DOI: http://dx.doi.org/10.1145/2470654.2466240

24. Cathal Gurrin, Gareth J. F. Jones, Hyowon Lee, Neil O'Hare, Alan F. Smeaton, and Noel Murphy. 2005. Mobile Access to Personal Digital Photograph Archives. In *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services (MobileHCI '05)*. ACM, New York, NY, USA, 311–314. DOI: http://dx.doi.org/10.1145/1085777.1085842

25. R. Harper, D. Randall, N. Smyth, C. Evans, L. Heledd, and R. Moore. 2008. The Past is a Different Place: They Do Things Differently There. In *Proceedings of the 7th ACM Conference on Designing Interactive Systems (DIS '08)*. ACM, New York, NY, USA, 271–280. DOI: http://dx.doi.org/10.1145/1394445.1394474

26. Steve Hodges, Lyndsay Williams, Emma Berry, Shahram Izadi, James Srinivasan, Alex Butler, Gavin Smyth, Narinder Kapur, and Ken Wood. 2006. SenseCam: A Retrospective Memory Aid. In *UbiComp 2006: Ubiquitous Computing*, Paul Dourish and Adrian Friday (Eds.). Lecture Notes in Computer Science, Vol. 4206. Springer Berlin Heidelberg, 177–193. DOI: http://dx.doi.org/10.1007/11853565_11

27. Benjamin Höferlin, Markus Höferlin, Daniel Weiskopf, and Gunther Heidemann. 2011. Information-based adaptive fast-forward for visual surveillance. *Multimedia Tools and Applications* 55, 1 (2011), 127–150. DOI: http://dx.doi.org/10.1007/s11042-010-0606-z

28. Nebojsa Jojic, Alessandro Perina, and Vittorio Murino. 2010. Structural epitome: A way to summarize ones visual experience. In *Advances in neural information processing systems*. 1027–1035. http://research.microsoft.com/apps/pubs/default.aspx?id=167024

29. Jacqueline Kerr, Simon J. Marshall, Suneeta Godbole, Jacqueline Chen, Amanda Legge, Aiden R. Doherty, Paul Kelly, Melody Oliver, Hannah M. Badland, and Charlie Foster. 2013. Using the SenseCam to Improve Classifications of Sedentary Behavior in Free-Living Settings. *American Journal of Preventive Medicine* 44, 3 (2013), 290 – 296. DOI: http://dx.doi.org/10.1016/j.amepre.2012.11.004

30. Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and Evaluation of a User Experience Questionnaire. In *HCI and Usability for Education and Work*, Andreas Holzinger (Ed.). Lecture Notes in Computer Science, Vol. 5298. Springer Berlin Heidelberg, 63–76. DOI: http://dx.doi.org/10.1007/978-3-540-89350-9_6

31. Hyowon Lee, Alan F Smeaton, Noel E OConnor, Gareth Jones, Michael Blighe, Daragh Byrne, Aiden Doherty, and Cathal Gurrin. 2008. Constructing a SenseCam visual diary as a media process. *Multimedia Systems* 14, 6 (2008), 341–349. DOI: http://dx.doi.org/10.1007/s00530-008-0129-x

32. Matthew L. Lee and Anind K. Dey. 2007. 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, New York, NY, USA, 131–138. DOI: http://dx.doi.org/10.1145/1296843.1296867

33. Matthew L. Lee and Anind K. Dey. 2008. 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, 44–53. DOI: http://dx.doi.org/10.1145/1409635.1409643

34. Cheng Li and Kris M. Kitani. 2013. Pixel-Level Hand Detection in Ego-centric Videos. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 3570–3577. DOI: http://dx.doi.org/10.1109/CVPR.2013.458

35. Siân E. Lindley, Richard Harper, Dave Randall, Maxine Glancy, and Nicola Smyth. 2009. 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, Article 2, 10 pages. DOI: http://dx.doi.org/10.1145/1613858.1613861

36. Catherine Loveday and Martin A Conway. 2011. Using SenseCam with an amnesic patient: Accessing inaccessible everyday memories. *Memory* 19, 7 (2011), 697–704. DOI: http://dx.doi.org/10.1080/09658211.2011.610803

37. Zheng Lu and Kristen Grauman. 2013. Story-Driven Summarization for Egocentric Video. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2714–2721. DOI: http://dx.doi.org/10.1109/CVPR.2013.350

38. Martina Luchetti and Angelina R Sutin. 2015. Measuring the phenomenology of autobiographical memory: A short form of the Memory Experiences Questionnaire. *Memory* ahead-of-print (2015), 1–11. DOI: http://dx.doi.org/10.1080/09658211.2015.1031679

39. Steve Mann. 1997. Wearable computing: A first step toward personal imaging. *Computer* 30, 2 (1997), 25–32. DOI: http://dx.doi.org/10.1109/2.566147

40. Gillian O'Loughlin, Sarah Jane Cullen, Adrian McGoldrick, Siobhan O'Connor, Richard Blain, Shane O'Malley, and Giles D Warrington. 2013. Using a Wearable Camera to Increase the Accuracy of Dietary Analysis. *American journal of preventive medicine* 44, 3 (2013), 297–301. DOI: http://dx.doi.org/10.1016/j.amepre.2012.11.007

41. Kadir Peker, Ajay Divakaran, and others. 2004. 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*, Vol. 3. IEEE, 2055–2058. DOI: http://dx.doi.org/10.1109/ICME.2004.1394669

42. Nemanja Petrovic, Nebojsa Jojic, and ThomasS. Huang. 2005. Adaptive Video Fast Forward. *Multimedia Tools and Applications* 26, 3 (2005), 327–344. DOI: http://dx.doi.org/10.1007/s11042-005-0895-9

43. Hamed Pirsiavash and Deva Ramanan. 2012. Detecting activities of daily living in first-person camera views. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2847–2854. DOI: http://dx.doi.org/10.1109/CVPR.2012.6248010

44. Corina Sas, Tomasz Fratczak, Matthew Rees, Hans Gellersen, Vaiva Kalnikaite, Alina Coman, and Kristina Höök. 2013. 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, New York, NY, USA, 1041–1046. DOI: http://dx.doi.org/10.1145/2468356.2468542

45. Abigail J. Sellen, Andrew Fogg, Mike Aitken, Steve Hodges, Carsten Rother, and Ken Wood. 2007. Do Life-logging Technologies Support Memory for the Past?: An Experimental Study Using Sensecam. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 81–90. DOI: http://dx.doi.org/10.1145/1240624.1240636

46. Ana R. Silva, Salomé Pinho, Luís M. Macedo, and Chris J. Moulin. 2013. Benefits of SenseCam Review on Neuropsychological Test Performance. *American Journal of Preventive Medicine* 44, 3 (2013), 302–307. DOI: http://dx.doi.org/10.1016/j.amepre.2012.11.005

47. Shingo Uchihashi, Jonathan Foote, Andreas Girgensohn, and John Boreczky. 1999. 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, 383–392. DOI: http://dx.doi.org/10.1145/319463.319654

48. Emma Woodberry, Georgina Browne, Steve Hodges, Peter Watson, Narinder Kapur, and Ken Woodberry. 2015. The use of a wearable camera improves autobiographical memory in patients with Alzheimer's disease. *Memory* 23, 3 (2015), 340–349. DOI: http://dx.doi.org/10.1080/09658211.2014.886703