Video-Recording Your Life: User Perception and Experiences

Daniel Buschek, Michael Spitzer, Florian Alt
Media Informatics Group, University of Munich (LMU)
Amalienstr. 17, 80333 Munich, Germany
{daniel.buschek, florian.alt}@ifi.lmu.de, spitzerm@cip.ifi.lmu.de

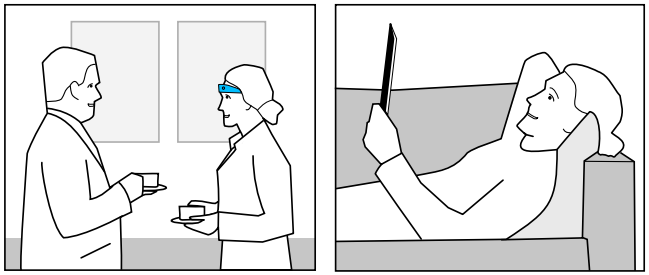

Figure 1: Wearable cameras allow us to record videos of our daily life (left, head-worn camera in blue). In combination with contextual information, short meaningful clips can be automatically created from the video footage. Users can review these clips at the end of their day (right), using them as digital human memory. This work explores the concept from the users' perspective, revealing strengths and challenges.

CHI'15 Extended Abstracts, Apr 18-23, 2015, Seoul, Republic of Korea
ACM 978-1-4503-3146-3/15/04.
http://dx.doi.org/10.1145/2702613.2732743

Abstract
Video recording is becoming an integral part of our daily activities: Action cams and wearable cameras allow us to capture scenes of our daily life effortlessly. This trend generates vast amounts of video material impossible to review manually. However, these recordings also contain a lot of information potentially interesting to the recording individual and to others. Such videos can provide a meaningful summary of the day, serving as a digital extension to the user's human memory. They might also be interesting to others as tutorials (e.g. how to change a flat tyre). As a first step towards this vision, we present a survey assessing the users' view and their video recording behavior. Findings were used to inform the design of a prototype based on off-the-shelf components, which allows users to create meaningful video clips of their daily activities in an automated manner by using their phone and any wearable camera. We conclude with a preliminary, qualitative study showing the feasibility and potential of the approach and sketch future research directions.

Author Keywords
Life logging; video recording; context; smartphone

ACM Classification Keywords
H.5.1 [Information interfaces and presentation]: Video.; I.2.10 [Vision and Scene Understanding]: Video analysis.

Scenario
Ellen recently joined a company building novel car UIs as a product manager. Today, she attends the annual sales meeting, seeing many colleagues from other cities for the first time. A lot of new ideas are discussed during the day. Back in her hotel room at night, Ellen replies to some emails, before watching the daily 120 s video-summary provided by her camera-augmented glasses. The video contains a discussion with one of her new colleagues. Despite not showing the full conversation, she immediately remembers that she wanted to send him the link to an article on this new 3D head-up display technology she recently read. What was his name again? Right, Jeff – smart guy. Together they could certainly build upon this new technology to create a novel 3D car navigation system.

Introduction

The advent of small, wearable cameras brings us closer to Vannevar Bush's vision of a computer-supported human memory [4]. Action cams, glasses, and smartphones become everyday companions to record our daily life. Today, this often happens in a selective manner – e.g. we record ourselves while doing outdoor sports – but in a not so distant future we may just leave the camera running throughout the whole day. These recordings can then be used, for example, to create a 120 second video summary of the day which helps us to recall people, events, activities and to-dos (cf. the scenario). Recordings could also serve as tutorials for others (e.g. for changing a flat tyre).

A major challenge for this vision is the creation of such video summaries from the vast amount of recorded material. Already today, a large proportion of recordings is never viewed or shared. Hence, we believe a mechanism is required to automatically identify and select interesting scenes, and to render a meaningful video clip. In this paper, we take a first step by investigating means for identifying interesting scenes in a video stream recorded in everyday life. We show that smartphone accelerometer data can be leveraged for scene selection. We evaluate our approach in a user study, where people were asked to record everyday life activities while simultaneously logging accelerometer data using a smartphone background app. We then create short video clips using our prototype and let users compare them to clips created by humans.

The contribution of this work is threefold: 1) We present findings from a survey about people's view on recording behavior. 2) We introduce a prototype that leverages accelerometer data to extract meaningful scenes from video streams. 3) An early qualitative study compares clips created with our prototype to manually cut videos.

Related Work

The idea of life-logging is almost as old as the computer itself. In 1945, Vannevar Bush envisioned the *Memex* – an electromechanical device to support and extend human memory by storing all knowledge we ever came in contact with for later access [4]. Despite never having been built, several concepts survived: In Microsoft's MyLifeBits project [2], Gordon Bell collects as much knowledge as possible from his life, using microphones or the SenseCam [6]. One aim of the project is to provide useful access to the gathered knowledge. While we share the project's vision of an approach accessible to the general public, we instead focus on an automated process with no inherent need for user interaction and using off-the-shelf hardware.

Work in the field of activity recognition assumes that points of interest correlate with people's activities. At the same time, prior projects show that multiple sensors are necessary to reliably detect activities (e.g., microphones [7], accelerometers [1], etc.). To make our approach work with one smartphone only, we opted not to focus on recognizing particular activities, but more fundamentally to detect changes in activity, following work from Blum et al. [3], who showed that changes in activity most likely occur out of interesting interruptions or new activities and may hence mark points of interest in themselves.

Survey

We conducted an online survey to assess user interest and behavior regarding video recording. The questionnaire had three parts: demographic information; creating, manipulating and sharing self-made videos; and attitudes towards creating video summaries of the user's day.

The questionnaire was created with LimeSurvey and distributed via mailing lists and Facebook. 57 people com-

pleted the questionnaire, mainly students and employees related to IT, media, and economics (24 male, mean age=25.7 years). They rated statements on a Likert scale (1=don't agree at all, 5=strongly agree).

Findings
Participants used different recording devices (multiple selections possible) – primarily ones with a small form factor. Most favorite devices were smartphones (41 participants), point-and-shoot cameras (16), DSLRs (13), webcams (7), tablets (5), and wearable cameras (3). Popular types of videos include videos of daily routine (29), holidays (29), and private events (26). 12 participants create videos while doing sports.

With regard to filming and processing behavior, we asked participants for each type of video 1) how generously they recorded scenes; 2) whether they cut scenes post-hoc; and 3) how much time they invest in post-processing. Across all types of videos, participants recorded rather much material (i.e. 4 on the Likert-Scale: daily routine 83.7%, private events 75.0%, sports 84.2%, holidays 80.0%).

The picture becomes more diverse regarding cutting and post-processing behaviour: Particularly for sports clips we found that 42.9% of our participants invest much or very much effort (4 and 5 on the Likert-Scale) in cutting and post-processing. For holiday videos, still 30.3% invest much or very much time in cutting. For all other types of videos, less than 25% stated to invest such effort in cutting and post-processing.

We also assessed sharing behavior. Here, we investigated whether users kept videos private, shared them with people they know, or made them publicly available. The vast majority did not at all like to share their videos with the general public (<6% for all types of videos). Sharing with friends was most popular (daily routine: 72.3%, private events: 72.2%, sports: 62.5%, holidays: 84.4%). The largest ratio of videos kept to themselves was observed for sports videos (31.2%).

Furthermore, we asked about quality requirements of 1) the raw video recording and 2) the final clip, depending on the intended audience. While quality requirements of raw recordings for unshared videos were high or very high for only 54% of participants (4 and 5 on the Likert-Scale), this was the case for 72% of participants for videos shared with friends, and for 85.6% regarding publicly shared videos. Findings for the final clips were similar with a slightly lower requirement for public clips (own use: 52.5%, friends: 72.5%, public: 75%).

With regard to the overall concept, participants expressed mixed views: While 24 participants (42.1%) could not imagine to use a system that creates video summaries of their day, 32 (56.1%) saw value. Being asked whether they would favor short clips or clips that prioritize important information, there was a strong tendency towards the latter version (37 participants, 71.1%).

Summary
The majority of users is not too selective when recording videos. They rather consider it important to capture all eventually important scenes. This suggests that users may indeed be happy to generously record video material throughout their daily life and implies the opportunity to automatically create more concise, meaningful clips.

Study
Our study gains insights into users' experiences with a system for automatic creation of life logging videos.

Apparatus

Our prototype consists of three parts (Figure 2): a GoPro wearable camera, a smartphone with acceleration logging app, and a data processing system on a desktop computer.

Processing Acceleration Data

We chose a simple feature for our prototype, extracting the change in average acceleration between subsequent parts of the video. We first compute average accelerations $\bar{x}, \bar{y}, \bar{z}$ in the i-th time frame ("window") W_i. The length of each window is a parameter of our algorithm, and subsequent windows overlap by 50%. This approach is in line with related work [5]. We then compute the acceleration differences between two subsequent windows. Finally, the total change δ_i is computed as the sum of the changes of the three dimensions.

Segmenting the Video

As a result, we derive a list of windows W_i with timestamps t_i, and associated acceleration changes δ_i. For this prototypical approach, we follow the basic assumption that higher acceleration changes indicate potentially more interesting events in the user's activities or context. This leads to the following segmentation process: For each timestamp, we add and subtract fixed durations d_s and d_e, to define a surrounding scene s_i with starting time t_i–d_s and ending time t_i–d_e. Both d_s and d_e are parameters of our algorithm; for the study we chose d_s=$2s$ and d_e=$5s$. Next, we merge scenes with overlapping time frames. Finally, we sort the resulting scenes s_i by their δ_i in descending order.

Selecting Scenes

Given a reduction factor η, our algorithm aims to reduce the video's total duration d to the new duration $d' = \eta d$. It selects scenes s_i from the top of the segmentation list, until their total length reaches d'. We chose $\eta = 0.1$ for this study. Finally, all unselected scenes are removed, resulting in an automatically created shortened clip.

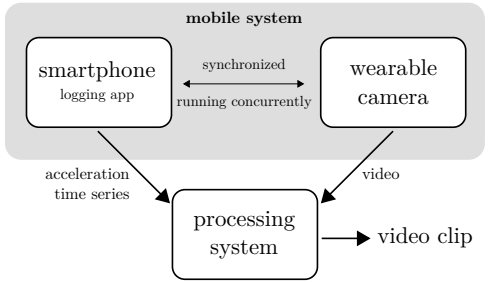

Figure 2: Prototype architecture with three parts: 1) a wearable camera, 2) a smartphone running an acceleration logging app, and 3) a data processing system. Camera and app are running synchronised during the user's day. The processing system analyzes the acceleration data post-hoc to cut the video accordingly and create a summarizing clip.

Procedure

We recruited 7 participants (19-29 years), all male students. They were compensated with a €30 gift card for an online shop. We also hired a hobbyist cutter, who received €80 for creating summaries. Thus, we wanted to simulate an intelligent human-like system able to identify important scenes and cutting them in a meaningful way.

Each participant was invited for an initial briefing, and asked to capture at least half an hour during the next day(s). Participants were free to decide what to record themselves. The only restriction was that they had to wear the camera and carry the phone. We did not inform participants about the specific purpose of the recordings. They were only informed that we would watch their videos, and that we would not distribute them further.

Participants returned to the lab after 1-2 days. We collected their recordings for further processing, and invited them to a final meeting a few days later. This gave our cutter time to create the manually composed videos.

Statement	Med.
I felt strange while recording.	2
I behaved differently than usual while recording.	2
I did something interesting while recording.	2

Table 1: User feedback regarding experiences during recording. (1: do not agree at all, 5: strongly agree)

In the final session, participants answered a short questionnaire about their experiences while recording (Table 1). Next, they were informed that we had created two summaries of their videos, and that we were interested in comparing two algorithms for doing so. Yet, we did not tell them that one video was created by a human cutter. They were shown both versions, answering a short questionnaire after each (Table 2). The order of the videos was counterbalanced. We revealed that a human cutter was involved at the very end of the study.

Quantitative Results
Participants answered a questionnaire about their experiences while recording (s) Table 1 summarises the results: Most participants felt not irritated by the camera, and used it in their everyday life. This fits the intended applications of our approach to life logging.

After watching each video, participants answered another questionnaire (Table 2). The algorithm was mostly rated neutral (3). A tendency towards too high cutting frequencies was perceived (algorithm's scenes: 7 seconds, cutter's scenes: 10 seconds). Overall, the manual cut was rated 1 (median) point better than the automated version.

The manual choice of cuts received the highest rating twice, and never the worst one, while our algorithm received the lowest rating twice, and never the best one. No one considered the algorithmic version not interesting at all, nor very interesting. However, the manual version also never received the highest rating here. These results indicate that automatic cutting with our approach is feasible, but not as good as manual scene selections.

For perceived quality of scene selections, the algorithm never received extreme ratings (1, 5). In contrast, human-cut videos received both the highest and lowest rating once. These results suggest that the cutter's personal choices may match or miss the user's taste, while the comparatively simple reasoning of our algorithm leads to more neutral selections as perceived by our participants.

Qualitative Results
The recordings from participant 2 and 3 contained few scenes showing others. Hence, the algorithm's cut contained a similar low ratio of such scenes. In contrast, the human cutter mostly selected exactly these scenes. This was perceived differently by our participants: Subject 3 agreed with the choice of the cutter, while subject 2 found these scenes awkward, since most people that this user had met had reacted more distant due to the camera. Including these reactions made the video worse in this user's point of view. On the other hand, our algorithm missed a scene of subject 7 having a quick chat with a friend, and of subject 5 stopping to watch river surfers.

These observations support the finding, that the algorithm results in more neutral scene selections than human reasoning. Participants' feedback suggests that this may be favourable, depending on other people's activities and reactions, which cannot be assessed by our prototype.

Discussion

Overall, the results from our studies show that users see value in the idea and that a basic prototype can create videos which lead to perception of key aspects and UX comparable to those of manually cut clips. Moreover, we collected valuable insights from the questionnaires and the qualitative feedback of the study regarding challenges and pitfalls of the approach. We believe our prototype to provide a valuable basis for future research that could center around the following aspects.

Question	Med.
Did you find the scene selection agreeable? (1: not at all, 5: very much)	4/3
Did you find the choice of cuts agreeable? (1: not at all, 5: very much)	4/3
Did you find the video to be interesting? (1: not at all, 5: very much)	4/3
Was the video exhausting to watch? (1: not at all, 5: very much)	1/2
What do you think about the cutting frequency? (1: too low, 5: too high)	3/4
What do you think about the video's length? (1: too short, 5: too long)	3/3

Table 2: User feedback questionnaire comparing manually cut videos (left) with automatically composed ones (right).

Most importantly, we found that people had different expectations regarding what should be included in the clip. For the manually cut videos this became apparent in cases where the cutter's taste did not closely match the participant's taste. As a result, future work should focus on how to meet these expectations by respecting (manually defined) user preferences. This is challenging since it may often be difficult to automatically determine scenes that match an individual's preferences. For example, if users prefer to include conversations (like subject 3), this can be achieved by analyzing audio data or via face detection. However, if the video contains many short conversations, (e.g. during a conference) it will be very difficult to algorithmically determine the most interesting ones. Solutions may include manually tagging particular scenes, for example by performing a subtle but easy-to-detect gesture in front of the camera [9].

Our results also suggest that length and cutting frequency should be more dynamic. Since participants perceived the video as neither too short nor too long, our chosen 180 seconds are a suitable length. However, users complained about particular scenes not being included. This suggests that it may be ok to exceed the fixed time limit on days with many interesting activities. Individual scenes were perceived as too short in the algorithm's videos, while users liked the variations and overall longer durations of scenes composed by the human cutter.

Conclusion

We have presented an approach for creating meaningful clips from video recordings of users' everyday life with the aim to serve as a digital memory. We showed that this is technically feasible by using accelerometer data from the user's smartphone only and that the approach is perceived as valuable by users. From our field study we obtained valuable hints for future research directions, such as a stronger focus on individual user preferences.

References

[1] Bao, L., and Intille, S. S. Activity recognition from user-annotated acceleration data. In *Pervasive computing*. Springer, 2004, 1–17.

[2] Bell, C. G., and Gemmell, J. *Total recall: How the e-memory revolution will change everything*. Dutton, 2009.

[3] Blum, M., Pentland, A. S., and Troster, G. Insense: Interest-based life logging. *IEEE MultiMedia 13*, 4 (Oct. 2006), 40–48.

[4] Bush, V., et al. As we may think. *The atlantic monthly 176*, 1 (1945), 101–108.

[5] Casale, P., Pujol, O., and Radeva, P. Human Activity Recognition from Accelerometer Data Using a Wearable Device. *Pattern Recognition and Image Analysis 6669* (2011), 289–296.

[6] Hodges, S., Williams, L., Berry, E., Izadi, S., Srinivasan, J., Butler, A., Smyth, G., Kapur, N., and Wood, K. Sensecam: A retrospective memory aid. In *Proc. of UbiComp '06*. Springer, 2006, 177–193.

[7] Ishiguro, Y., Mujibiya, A., Miyaki, T., and Rekimoto, J. Aided eyes: eye activity sensing for daily life. In *Proceedings of the 1st Augmented Human International Conference*, ACM (2010), 25.

[8] Khan, A. M., Lee, Y.-K., Lee, S., and Kim, T.-S. Human activity recognition via an accelerometer-enabled-smartphone using kernel discriminant analysis. In *Future Information Technology (FutureTech), 2010 5th International Conference on*, IEEE (2010), 1–6.

[9] Knibbe, J., Seah, S., and Fraser, M. VideoHandles: Replicating Gestures to Search through Action-Camera Video. In *Proceedings of the 2nd ACM Symposium on Spatial User Interaction* (2014), 50–53.