The Value of Information Cues for Lifelog Video Navigation

Katrin Wolf[1], Lars Lischke[2], Corina Sas[3] and Albrecht Schmidt[2]
[1]Hamburg University of Applied Science, [2]University of Stuttgart, [3] Lancaster University
[1]katrin.wolf@acm.org, [2]firstname.lastname@vis.uni-stuttgart.de, [3]c.sas@lancaster.ac.uk

ABSTRACT
With the advent of lifelogging cameras the amount of personal video material is massively growing to an extent that easily overwhelms the user. To efficiently review lifelog data, we need well designed video navigation tools. In this paper, we analyze which cues are most beneficial for lifelog video navigation. We show that the information kind determines the most appropriate cue in single cue systems, but that multicue approaches are more appreciated. These findings can inspire to design video players with multiple navigation cues, including time, place, persons, and events for easier and more efficient lifelog video retrieval.

ACM Classification Keywords
H.5.2 User Interfaces: Graphical User Interface.

Author Keywords
Video navigation; lifelog video; video retrieval.

INTRODUCTION AND BACKGROUND

Due to the ubiquitous availability of cameras, such as those embedded into phones or through lifelogging devices like GoPro cameras or Google Glass, private video collections are massively growing. Surprisingly neither design guidelines nor tools exist that support an efficient and easy retrieval of lifelog video. Hence, research on video navigation is crucial to gain basic knowledge about navigation aids users can benefit from when retrieving specific video scenes.

We believe that personal video navigation should as similar as possible be designed in a way humans cognitively access their autobiographical memory. It is widely accepted that autobiographical memories of past situations are cognitively accessed by various memory cues, such as what happened, who was there, when it took place, and where it occurred [1, 3, 18]. It had been observed that "what" happened is one of the most important information that people want to recall to memorize a certain life event. It is also known that providing additional information about with "who", "where", and "when" the "what" was experienced helps a lot to recall a situation. Moreover, multiple cues are more effective than single cues for recalling a situation [18], e.g., recalling what happened is greatly improved by providing both "who" and "where" compared with providing just one or the other. However, the exact importance of different cues may depend upon the specificity of the cue and the way in which the life events are organized, such that "who" can be the most effective recall cue if the recalled person is very important to the recalling person [3].

Analyzing existing video navigation approaches, e.g. YouTube, shows that players organize video through a time line, but they do not display the time at that video was recorded even though this is embedded in video meta-data. Moreover, the location at that the video was recorded if often saved as GPS information, but again, that information is not presented when watching the video with state-of-the-art players. Finally, faces of persons shown in the video could be added manually as labels or be detected with image processing, which is already used for organizing photos, like with the iPhoto software. Research on multimodal video annotation provides aid for automatically detecting place [13] and person [8]. However, a autobiographical founded multicue approach has not been investigated. Hence, considering the memory cues time, place, and person for personal video presentation and navigation is a true research gap.

Researchers explored alternative ways of video navigation. Examples are different time-based video arrangements, such as a clock-based arrangement of video frame thumb nails [10] or slit scans reducing a frame, for example, to a pixel column and creating slit scan frames out of several of these pixel columns [14]. 2D storyboard arrangement were used to give an overview of the video content via thumb nails [2, 4, 6, 11, 17]. Storyboards enable for displaying importance of scene through re-sizing specific thumb nails, e.g., those that were most often watched. In all mentioned time-based navigation types, video is represented according time information, usually in the sequence the data was captured. However, these approaches often do not provide information about absolute times of data recording. Chiu et al. [5] used 3D storyboards to create a virtual 3D city in which the user can browse through different videos. Here every building displays one video, and the facades and roofs contain storyboards with key frames. Location-based approaches with videos placed on a map allow for video navigating through the information about the place it was recorded [9, 12, 15]. Christel [6] used multiple meta-information for video indexing, like location, time, and semantic text labels. Then, one representation type per index information was created, which results in multiple video navigation systems for the same video content: a storyboard for

MUM'16, December 12–December 15, 2016, Rovaniemi, Finland.

http://dx.doi.org/10.1145/3012709.3012712

time information, a map for the location information, and a mind map for text based labels. The existing body of research in video navigation is mainly using single navigation cues, such as time, place, and person. Only some hybrid systems exist. However, to date neither a systematic investigation on the navigation value of the cues time, place, and person has been conducted nor the benefit of providing all three cues for video navigation has been explored.

This paper extends previous research through (1) a systematic investigation of the value of the navigation cues time, place, and person and (2) an exploration of systems that combine these three cues in one single video player.

METHOD

We investigate the benefit of the navigation cues person ("who"), place ("where"), and time ("when") using 2h40 quasi-lifelog video recorded by us to ensure that every participant solved the tasks under the same difficulty. Of course, 2h40 hours barely represent lifelog video. However, we believe that a lifelog video navigation system would require to segment the data into smaller chunks (like movie scenes). Our short video represents such chunks. We are aware that using somebody else's lifelog video is artificial. However, using the same video material for all participants has the advantage to guarantee equal difficult tasks for all participants. Moreover, personal videos contain emotional meta-information that would influence our results as stated by Wagenaar: "Pleasant events were better recalled than unpleasant events" [18]. To avoid that, we use video participants have no emotional connection with. To ensure that our participants were familiar with the video as if it was their own data, we provided intense video content training and reminders before and during the tasks. The experiment was divided into two parts. In the *single cue part* we focused on effects of each cue and hence, we isolated them resulting a three single-cue video browsers, one supporting time, one persons, and one location. In the *multicue part*, we explored the benefit of multicue navigation. We tested two hybrid video browsers, once with a strong time-based arrangement and one using a location-based placements of the video sequences as time- and location-based designs are already used in common video and image navigation systems, like Youtube and Google, while person-based video navigation has not been proposed yet.

Apparatus

As apparatus we implemented 5 different video players in Unity. In each player the same video content was presented. The video was continuously captured with a head-worn GoPro with 30 fps. Each player represented one condition, 3 used in *single cue part* of our experiment and 2 in *multicue part*, which are described as follow:

Person-based player: Here the video is represented according the persons shown in the video, see the left image in the top of Figure 1. After clicking on a person's icon, the player shows the videos on that the person is shown the first time. The person's icon size corresponds with his screen time to indicate how long one person is shown in the video compared to another. All video sequences of one person are stitched together in the order they were recorded. As one person was always with the person who wore the camera, one can access the entire video through clicking on this person's icon. The video can be controlled with a *PLAY* and *PAUSE* button, with the mouse through grabbing the player handle, and alternatively for fine forward and back jumps and to pause via the navigation keys *Delete* to skip backwards, with *End* to pause the video, and through *Page Down* to skip forward. The time-, person, and locations-based players save task completion time (TCT) in logfiles.

Location-based player: In this player, see the middle top image of Figure 1, the video is arranged according the location it was recorded. Consequently, the video is sliced in short sequences and placed on a map. If the camera-wearing person was in motion, e.g. walking or going by train, the video is represented in video slices on a path across the map. These small videos store 10 seconds of video each. If the camera wearing person is longer at one location, the video stores the entire material captured there, and the video player has a larger size. Zooming in/out in the map is possible using the mouse wheel. Drag and drop actions allow for panning the map. The space key changes the perspective back to the initial bird view. The video players one the map are icons, and clicking on them results in zooming in to see the players from the front and for playing back the video. The videos can again be controlled using the similar keys as in person-based players.

Time-based player: In this player, see the right top image of Figure 1, video is represented according time. The player shows the entire video and labels it with real time stamps. The player control is same like for the person-based player.

Time-dominated multicue player: This player is designed as a combination of the person- and the time-based players, see Figure 1 on the left bottom. Thus, the faces of person can be selected to access the sequences where certain persons were met. Moreover, the player integrates time and date. In addition a 2D map shows where the current video sequence was recorded. The functions of the player are designed similarly to the time- and person-based player versions.

Location-dominated multicue player: For this player, shown in Figure 1 on the right bottom, the 3D map is chosen as base, and the navigation aid of persons shown in the video can be accessed from the bird view through clicking on face icons. In the perspective when moving across the map for watching the video sequences, information about time and date are shown. The player control is similar to the location-based version.

Task

In *single cue part*, for each of the three video players (providing cues about times, places, and persons) the participants were asked to solve the same nine tasks. Three tasks required to search for a specific time, like: "Navigate to the video frame at 3:10pm shown by a clock in the video". The three tasks focusing on a location asked to roughly navigate to a frame when a certain place was entered, like: "Go to the video frame when the person that wears the camera enters university". The three tasks focusing on a person, asked to browse

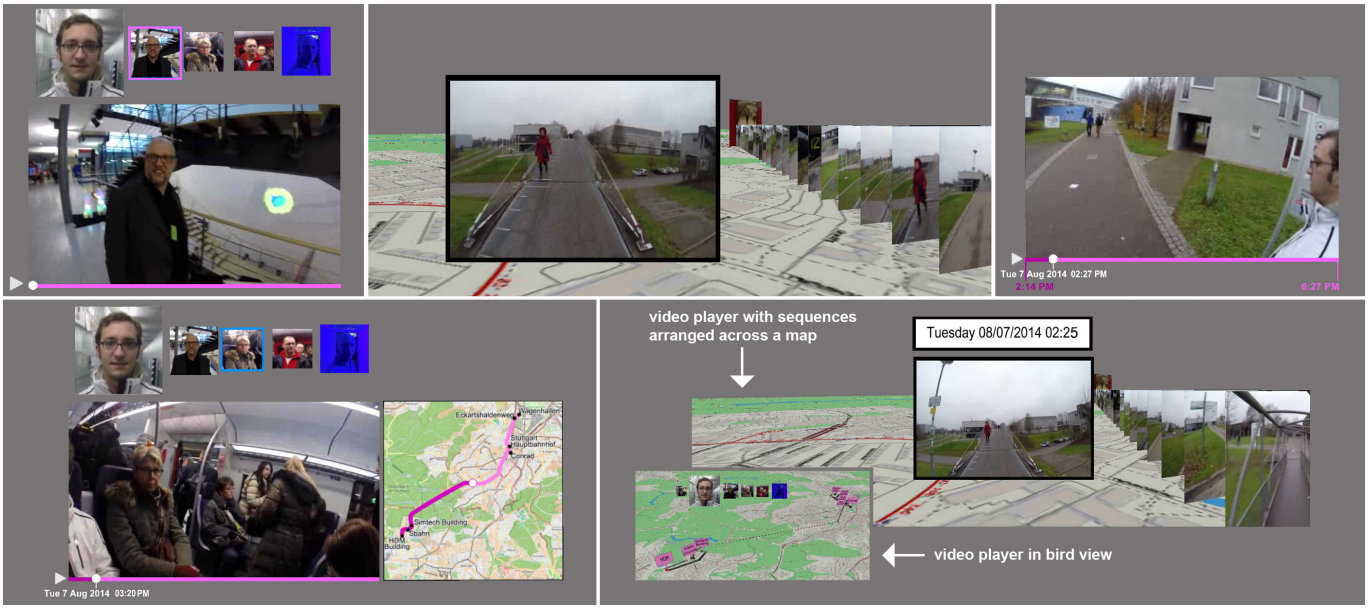

Figure 1. Top left: Video player with person as navigation cue. Top center: Video player with location as navigation cue. Top right: Video player with time as navigation cue. Bottom left: Video player integrating all cues, but having time as dominant cue. Bottom right: Video player integrating all cues, but having location as dominant cue.

to a frame where a certain person was met, like: "Go to the frame when the person that wears the camera meets Niels the first time during the video". Because of the focus on lifelog videos, it is suitable to use the same video content and questions for all conditions, because we assume the user roughly knows his or her own lifelog video, which is ensured through training. The tasks were solved through roughly browsing to the right frame, pause the video, press a *DONE* button, and confirm the selection. If the *DONE* button was accidentally pressed, the task could be continued through rejecting the confirmation. If participants were not able to solve a task, they could skip it through pressing a dedicated button.

In *multicue part*, we asked the participants to separately explore our 2 hybrid video browsers for about 2 minutes each until they had understood the underlying navigation concept. As we combined the cues in the hybrid systems, quantitative data analysis would not provide us with insight about the benefits of the design or isolated cues. Here, we rather aimed to deepen the understanding of multi-cue navigation. Hence, we decided to gather and to analyze qualitative data.

Procedure

After the participants were welcomed and informed about the study purpose, they signed a consent form and filled in a demographic questionnaire. As we aimed to measure navigation time for personal lifelog video, which means for video that the user roughly knows, we provided the following support for the navigation tasks in *single cue part*: We showed them a 10 min video summary of the 2h40 lifelog video, which was played back in real time when the video showed the answers to the tasks and in 25 times faster speed during all other sequences. To measure navigation performance, but not the ability to remember video content, participants got during the tasks a sheet with a brief summary of the video content as well as with the screen shots of situations they should navigate to. They also got a map of our city where the locations they had to search for were marked. Before every condition participant got a training and were introduced to the specific video player controls. After each condition, participants rated their perceived mental load on the SMEQ scale [19] and gave qualitative feedback about aspects that helped during the navigation tasks or about functions that they had missed. During *multicue part* participants first got a training into the controls of both, the time-dominated and the location-dominated hybrid player. After the training participants explored the player. Then they gave qualitative feedback through open questions about beneficial aspects and missed functions of the player. To compensate their time, the participants got candies.

Design

In the experiment 18 participants (8 females, ages from 21-43, M=26.4, SD=5.3) took part. As described, the study was divided into two parts. *single cue part* had a within subjects design with the independent variable video navigation cue (time, place, person). For that part, the dependent variables were task completion time (TCT) measure automatically through Unity and saved in logfiles, perceived mental effort rated on the SMEQ scale [19] as it is known to be very sensitive with small sample sizes [16], and qualitative feedback that participants typed into an online form. Using a repeated measures design, every participant performed all 9 navigation tasks with all three video players as conditions. The order of the conditions was counterbalanced. The order of the tasks within each condition was randomized. *multicue part* again had a within subject design with the independent variable video player design (time-dominated multicue player, location-dominated multicue player). This time only qualitative feedback was recorded via open questions. The condition order was counterbalanced.

		Person-based		Location-based		Time-Based	
		M	*SD*	*M*	*SD*	*M*	*SD*
Task	Person	7.7	3.6	48.7	12.8	23.3	11.4
	Location	24.8	11.4	12.8	5.2	19.6	7.4
	Time	39.4	16.8	39.0	30.4	22.6	13.1

Table 1. TCT means & SD for browser type * task in sec.

Person-based		Location-based		Time-Based	
M	*SD*	*M*	*SD*	*M*	*SD*
23.9	17.5	33.5	24.4	21.8	10.8
Person		Location		Time	
M	*SD*	*M*	*SD*	*M*	*SD*
26.6	19.7	19.0	9.6	33.7	22.5

Table 2. TCT means & SD for player (top) & task (bottom) in sec.

RESULTS

We first present the quantitative analysis of the data recorded in *single cue part*, followed by the qualitative comments collected in both *parts*.

Perceived effort: For the effort perceived measured during *single cue part*, we conducted a one-way ANOVA and found a significant effect of the player used on the effort ratings ($F_{2,34}$=4.526, p=.018, time M=16.4 & SD=16.5, person M=28.5 & SD=23.5, location M=32.3 & SD=26.8). While post-hoc tests using Sidak corrected comparisons showed that the location-based player was perceived to be significantly harder to use than the time-based browser (p=.039), neither a difference in effort was perceived between using the time- and person-based player (p=.074) nor between the location-based and the person-based design (p=.890).

Task completion time: Three-way ANOVAs yielded significant effects of condition and task on TCT (condition: $F_{2,88}$=8.087, p=.001, task: $F_{2,106}$=15.793, p<.001) as well as a significant interaction effect between both factors ($F_{4,120}$=23.078, p<.001), see Table 1. Sidak corrected post-hoc tests showed that the location-based player took significantly longer than the other conditions (location vs. time: $p < .001$, location vs. person: p=.011), but person- versus time-based design did not show a significant difference in TCT (p=1.000). Regarding task, post-hoc tests indicated finding locations was the longest task (location vs. time: p=.006, location vs. person: p=.006), and that searching for persons took the participants longer than searching for times (p=.018), as also shown in Table 2.

Time, place, and person cues: Unsurprisingly the cues time, place, and person were most appreciated when searching for video sequences with similar content. The location-based player was appreciated most when searching for places. Consequently, when searching for an information that was not represented by the player's navigation cue, participants often missed the corresponding cue. Alternative navigation strategies, if the right navigation cue was missing, were the chronic description of the events that happened in the video (representing one's memory) in combination with information about "where" something happened. That enabled the participants to then search for the "who" or the "when" in the scene. When searching for time using the locations-based player, one participant mentioned: "I tried to remember the locations that mapped the time". An additional navigation cue that participants missed relates to the "what" that happened. One participant, for example, missed in the map "labels for events, such as meeting persons" and another participant proposed to "split videos when meeting an important person for the first time", which suggests to introduce scenes in lifelog video such as we are used to from movies. That was proposed to provide video "bookmarks", e.g. for "skipping sections" or for "jumping to the first occurrence of a person".

Over all, participants appreciated both, the location-and the time-dominated multicue design. One participant said about the location-dominated version: "Absolutely nice, very fast search tool", and another said about the time-dominated browser: "Of cause the real time is the most useful feature, but also the mini-map and faces. Everything was well displayed." However, some little improvements were suggested, like a map to click on for the time version and the possibility to jump to certain times for the video map.

DISCUSSION & CONCLUSION

Our results show that each autobiographical memory cue implemented as navigation cue in lifelog video players helps to find corresponding information. To browse to a specific moment in time, to a person or to a location, the corresponding design enables fastest navigation. Users had overall the shortest search time with the time-based player. This might be explained by the fact that time-based video navigation is the common player approach. However, the map-based version, even though being slower, was appreciated a lot when searching for locations but also for persons, as users find people in video through remembering the place they were met and can then from such starting point precise their search. The higher perceived effort while searching particular sequences with the location-based player might be due to the unknown navigation concept and unfamiliar visualization of the video. Furthermore, one might perceive the 3D representation to be overwhelming as 3D user interfaces have been found to be less efficient [7]. This is also supported by the positive feedback regarding the time-dominated hybrid video player, where we used a 2D map to display location.

We conclude that a multicue solution allows for retrieving multiple information kinds and is favoured by participants. Multicues provide a lifelog video player with navigation cues that represent the main memory cues, which we found to be a promising direction for future lifelog video players. In addition events are desired as cues, and we suggest (according to [18]) to consider, beyond the "who", the "when", and the "where", also the "what" when designing lifelog browsers.

In summary, we contribute to the field of video lifelogging through (1) showing that autobiographical memory cues can beneficially be used as video navigation cues (2) the cue type fastest retrieves information of the same kind (3) a multicue solution would allow for retrieving multiple kinds of information, is favored by participants, and hence, is a promising approach for future lifelog video navigation systems.

REFERENCES

1. Lawrence W Barsalou. 1988. The content and organization of autobiographical memories. *Remembering reconsidered: Ecological and traditional approaches to the study of memory* (1988), 193–243.
2. 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
3. Christopher DB Burt. 1992. Retrieval characteristics of autobiographical memories: Event and date information. *Applied Cognitive Psychology* 6, 5 (1992), 389–404.
4. 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
5. Patrick Chiu, Andreas Girgensohn, Surapong Lertsithichai, Wolf Polak, and Frank Shipman. 2005. MediaMetro: Browsing Multimedia Document Collections with a 3D City Metaphor. In *Proceedings of the 13th Annual ACM International Conference on Multimedia (MULTIMEDIA '05)*. ACM, New York, NY, USA, 213–214. DOI: http://dx.doi.org/10.1145/1101149.1101182
6. Michael G. Christel. 2008. Supporting Video Library Exploratory Search: When Storyboards Are Not Enough. In *Proceedings of the 2008 International Conference on Content-based Image and Video Retrieval (CIVR '08)*. ACM, New York, NY, USA, 447–456. DOI: http://dx.doi.org/10.1145/1386352.1386410
7. Andy Cockburn and Bruce McKenzie. 2002. Evaluating the Effectiveness of Spatial Memory in 2D and 3D Physical and Virtual Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '02)*. ACM, New York, NY, USA, 203–210. DOI: http://dx.doi.org/10.1145/503376.503413
8. Adam Fouse, Nadir Weibel, Edwin Hutchins, and James D. Hollan. 2011. ChronoViz: A System for Supporting Navigation of Time-coded Data. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems (CHI EA '11)*. ACM, New York, NY, USA, 299–304. DOI: http://dx.doi.org/10.1145/1979742.1979706
9. 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
10. Mieke Haesen, Jan Meskens, Kris Luyten, Karin Coninx, Jan Hendrik Becker, Tinne Tuytelaars, Gert-Jan Poulisse, Marie-Francine Moens, and others. 2013. Finding a needle in a haystack: an interactive video archive explorer for professional video searchers. *Multimedia tools and applications* 63, 2 (2013), 331–356.
11. Dan Jackson, James Nicholson, Gerrit Stoeckigt, Rebecca Wrobel, Anja Thieme, and Patrick Olivier. 2013. Panopticon: A Parallel Video Overview System. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. ACM, New York, NY, USA, 123–130. DOI: http://dx.doi.org/10.1145/2501988.2502038
12. He Ma, Roger Zimmermann, and Seon Ho Kim. 2012. HUGVid: Handling, Indexing and Querying of Uncertain Geo-tagged Videos. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '12)*. ACM, New York, NY, USA, 319–328. DOI: http://dx.doi.org/10.1145/2424321.2424362
13. Alistair Morrison, Paul Tennent, John Williamson, and Matthew Chalmers. 2007. Using Location, Bearing and Motion Data to Filter Video and System Logs. In *Proceedings of the 5th International Conference on Pervasive Computing (PERVASIVE'07)*. Springer-Verlag, Berlin, Heidelberg, 109–126. http://dl.acm.org/citation.cfm?id=1758156.1758165
14. Michael Nunes, Saul Greenberg, Sheelagh Carpendale, and Carl Gutwin. 2006. Timeline: Video traces for awareness. In *Video Proc. ACM CSCW*, Vol. 6.
15. Suporn Pongnumkul, Jue Wang, and Michael Cohen. 2008. Creating Map-based Storyboards for Browsing Tour Videos. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology (UIST '08)*. ACM, New York, NY, USA, 13–22. DOI: http://dx.doi.org/10.1145/1449715.1449720
16. Jeff Sauro and Joseph S. Dumas. 2009. Comparison of Three One-question, Post-task Usability Questionnaires. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 1599–1608. DOI: http://dx.doi.org/10.1145/1518701.1518946
17. 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
18. Willem A Wagenaar. 1986. My memory: A study of autobiographical memory over six years. *Cognitive psychology* 18, 2 (1986), 225–252.
19. FRH Zijlstra and L van Doorn. 1985. The construction of a scale to measure subjective effort. *Delft, Netherlands* (1985).