
Mobile Interaction Techniques for Interrelated Videos

Jochen Huber

Technische Universität Darmstadt
Hochschulstraße 10
Darmstadt, 64289 Germany
jhuber@tk.informatik.tu-darmstadt.de

Jürgen Steimle

Technische Universität Darmstadt
Hochschulstraße 10
Darmstadt, 64289 Germany
steimle@tk.informatik.tu-darmstadt.de

Max Mühlhäuser

Technische Universität Darmstadt
Hochschulstraße 10
Darmstadt, 64289 Germany
max@tk.informatik.tu-darmstadt.de

Abstract

With the advent of increasingly powerful mobile devices like Apple's iPhone, videos can be used virtually anywhere and anytime. However, state of the art mobile video browsers do not efficiently support users in browsing within individual, semantically segmented videos and between the large amounts of related videos, e.g. available on the Web. We contribute a novel user interface for the mobile navigation of large video collections comprising two novel spatial interaction techniques for the mobile, nonlinear interaction with multiple videos. Evaluation results show that our solution leads to significantly higher efficiency and user satisfaction.

Keywords

Multimedia, mobile devices, video browsing, e-lectures

ACM Classification Keywords

H5.2. Information interfaces and presentation: User Interfaces.

General Terms

Design, Human Factors

Copyright is held by the author/owner(s).
CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.
ACM 978-1-60558-930-5/10/04.



Figure 1. iTunes U digital library browser on the iPhone OS. Users can either search for lectures using the search button at the bottom or select appropriate lectures from various categories. However, lectures can only be watched as an ordinary movie. It is impossible to either access slides directly or navigate between semantically related e-lectures.

Introduction

Increasingly powerful mobile devices like Apple's iPhone continuously shape how we perceive multimedia when being on the move. Users are able to access millions of video streams e.g. through the iTunes store almost anytime and anywhere. Browsing of individual videos on mobile devices has been addressed by only a few research projects: most notably (1) MobileZoomSlider [4], allowing users to skim through individual video streams quickly by adapting the playback speed through a rubber-band metaphor and (2) MiniMedia surfer [6], supporting keyword queries and allowing users to explore query results through key frames. The latter browser's navigation completely relies on the designated keywords. This is a major issue when e.g. trying to get an overview without knowing exactly what to look for. Besides watching single videos, the interrelationship of this video data (e.g. as hyperlinks in so-called *hypervideos*) is of major importance, analogously to for instance textbooks and their contained references. However, current mobile video browsers do neither support the efficient navigation within single, semantically segmented videos, nor the navigation between multiple, e.g. topically overlapping videos.

The above observations let us formulate two key requirements for mobile video browsers:

1. Mobile video browsers shall not only support users when watching a *single* video but shall highlight the very relationships *between various* videos. Hence, users will be able to browse an interwoven web of videos. As a direct consequence, browsers shall support users in *getting an overview on* and navigating *between* topically overlapping videos.

2. Users shall be able to use this interwoven web of videos efficiently on mobile devices, overcoming their limited device characteristics like small form factors and displays.

Based on these requirements, we have developed an interface concept for the mobile navigation of large, semantically interrelated video libraries, which is to the best of our knowledge the first approach. This comprises two novel, spatial interaction techniques for the mobile, nonlinear interaction with videos. This is relevant for various fields. In the following, we outline two examples.

1. Recordings of Talks and Lectures

E-Lectures consist of various, synchronous multimedia streams, typically an audio recording of the lecturer's talk (*audio stream*) and (probably annotated) presentation slides (*whiteboard stream*). The streams can be semantically segmented using the lecture slides as key frames, which each represent a semantic unit.

The ubiquitous availability of multimedia learning material through services like iTunes U [5] or OpenCourseWare [7] has paved the way for groundbreaking changes in mobile learning. A recent study [3] found a shift in the usage habits of students towards using the mobile version of e-lectures. Fostering a good learning process should not only comprise the usage of individual e-lectures. Various topically related lectures from different institutes allow learners to for instance receive elaborate explanations for a certain problem.

However, state of the art mobile video browsers do not support the user sufficiently in these tasks, which involve the use of *multiple* e-lectures. A learner would

have to (1) identify potential lectures in the digital library browser (see Fig. 1), (2) scan each lecture sequentially to check whether it really covers the right topic and (3) note down or memorize the occurrences and correct positions within the e-lecture. Hence, it is impossible for learners to complete this task in a reasonable amount of time in a mobile setting.

The arrow up indicates that other videos contain content, which is similar to the currently viewed video. By flicking downwards, the user interface moves up and reveals the related key frames (see Fig. 3).

Users can navigate directly through the video by flicking horizontally. Flicking from right to left switches to the next key frame and vice versa.

The key frames are aligned in a grid. Key frames can be switched by simply tapping onto the miniaturized versions. It is also possible to zoom in by double tapping the grid. The current key frame is highlighted in the grid.

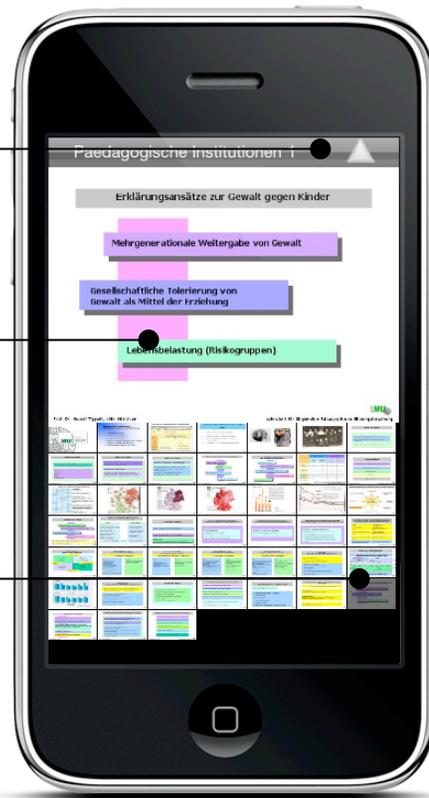


Figure 2. Screenshot of our mobile video browser.

2. Ethnography

The abilities of mobile devices enable ethnographers to examine details of action during observations and therefore open new opportunities for digitally augmented environments (e.g. with various video cameras in driving cars) [2]. Typically, the recorded data is being manually or automatically analyzed and annotated. Moreover, it is semantically segmented through e.g. key frame abstraction. A key frame would for instance correspond to an observation sequence as a semantic unit. Last, the data is being interrelated (e.g. temporally and logically). Although, these novel devices allow researchers to record the data in the field, they lack support for browsing the *preprocessed* data efficiently while being on the move.

In the remainder of this paper, we first present our concept before reporting on evaluation results. Finally, we discuss our findings and point out potential future work.

Concept

The main goal for our interface concept can be deduced from the requirements for mobile video browsers formulated above. The interface shall allow for an intuitive interaction *within* and *between* videos. Moreover, it shall foster awareness of video interrelationships, despite the mobile device's small screen. Due to these facts, we have utilized a simple but powerful spatial, two-dimensional metaphor. The horizontal dimension is used to browse within a semantically segmented video. The vertical dimension is used to navigate between topically related videos. Topical relationships are expressed as hyperlinks between the semantic units, here: key frames. We assume that these links are created automatically, e.g. through multimedia informa-

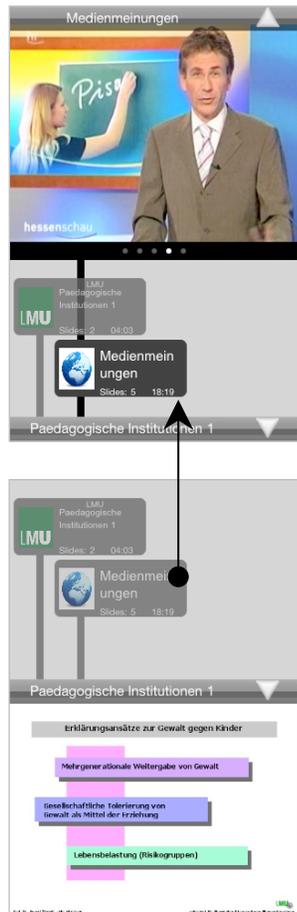


Figure 3. Browsing topically related videos.

tion retrieval [1]. Furthermore, it is also possible to enhance the user interface to allow users to manually create links between key frames.

Horizontal Navigation: Within a Video

Figure 2 shows a user interface screenshot of our video browser (here: an e-lecture with slides as key frames). The user interface is subdivided into two areas. The upper part shows the current video in detail, whereas the lower part shows an overview over all key frames. Users can navigate through the key frames by simply flicking horizontally through the upper part of the user interface. Either rotating the device into landscape mode or double tapping the current video in the upper part can start playback of the video. When playing the video in landscape mode, users can also navigate through the key frames by simply flicking horizontally.

Vertical Navigation: Between Videos

Whenever a video overlaps topically with other videos in the video collection (e.g. two keyframes cover the same topic), available relationships are indicated by a small arrow in the upper right corner of the user interface (see Fig. 2). When the user flicks downwards, the interface is being scrolled downwards, revealing related videos as shown at the bottom of Figure 3. In this case, two interlinked videos (visualized using grey boxes) contain relevant material. By tapping on one of the videos, the interface is being scrolled down further, thereby displaying the interlinked key frames of the related video (see Fig. 3 top). In turn, these can also contain topical relations to other videos, which are thence visualized again with a small arrow in the upper right corner.

By aligning semantically related videos vertically, the browsing history results in a vertical stack. This can be navigated by simply flicking vertically up and down respectively. Alternatively, to avoid repetitive flicking and to gain an overview on the browsing history, a visualization thereof can also be used for the vertical navigation as shown in Figure 4.

Evaluation

We have implemented the concept as part of an e-lecture browser for the Apple iPhone. It has been evaluated in a controlled experiment with 44 participants (30 male, 14 female) with different backgrounds. Each single-user session lasted about 2 hours. The overall goal was to evaluate the effectiveness, efficiency, learnability of the video browser, as well as user satisfaction. The tasks of the participants comprised simple fact-finding tasks, as well as advanced knowledge integration tasks. As data, we utilized e-lectures of each about 90 minutes length and the corresponding slides as key frames. Both, time required to complete the tasks and usability errors were measured. For each task, a different set of e-lectures was utilized to exclude any learning effects. The sessions were video-recorded and semi-structured interviews were conducted.

The experiment was subdivided into two parts (within-subject). The first one concentrated on navigating within single, semantically segmented videos (*intra-video navigation*) using the horizontal dimension. The second part focused on the navigation of interrelated videos (*inter-video navigation*), therefore adding the vertical dimension. This subdivision allowed us to assess the specific influence of each dimension on the



Figure 4. The vertical navigation history is displayed as an image on top of the current video (highlighted underneath the finger). By moving the finger vertically across the images, users are able to navigate forth and back within the browsing history.

usability and user experience goals. In the following, we report on the evaluation results for both parts.

Intra-Video Navigation

The participants were presented three different user interfaces: (1) an enhanced standard iPhone media player as baseline (*Baseline* in Fig. 5), providing additional buttons to switch back and forth between slides, (2) a player, which allows users to flick through the slides horizontally in landscape (*Flick* in Fig. 5) as with our video browser when rotated in landscape mode and (3) our video browser as shown in Fig. 2 without the possibility of inter-video navigation (*Grid* in Fig. 5). We introduced the flicking-based player to particularly distinguish this concept from the baseline on the one hand. On the other hand, we wanted to investigate the specific influence of the grid, when comparing our grid-based to the flicking-based concept.

The participants were asked to complete three different fact-finding tasks with each user interface. Task 1: the participants had to search an e-lecture for a given slide

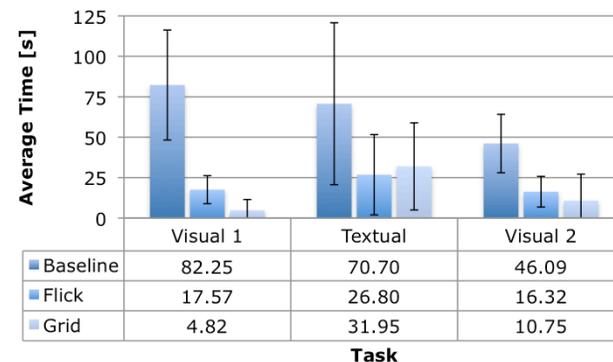


Figure 5. Average Time per Task for Intra-Video Navigation

without prior knowledge of the lecture (*Visual 1* in Fig. 5). Task 2: the participants were asked to find a certain topic in the last third of the lecture (*Textual* in Fig. 5). Task 3: the participants had to navigate to the slide following the one found in the first task (*Visual 2* in Fig. 5). Figure 5 shows an overview of the average required time per task and user interface. The participants were able to complete all three tasks significantly faster ($p < 0.001$) using either the flicking-based or the grid-based video browser than using the baseline player. The participants also committed about 60% less usability errors (significant with $p < 0.01$).

Comparing the grid-based with the flicking-based browser, we found that the participants were significantly faster using the grid-based browser for task 1 ($p < 0.001$) and task 3 (Fig 5., $p < 0.05$). In task 2, the difference was not significant. This is inline with qualitative findings from the semi-structured interviews. The participants stated that the grid-based browser supports their visual orientation and navigation (as in task 1 and 3), whereas they prefer to flick through the slides when they have no visual clues (as in task 2).

Inter-Video Navigation

The participants were presented two different user interfaces. First, a further enhanced standard iPhone media player (*Baseline* in Fig. 6), which allows switching forth and back between slides. Moreover, it displays topical relationships as textual hyperlinks on the slides. Second, the participants were asked to utilize our video browser with both, horizontal and vertical navigation capabilities (*2D-Nav* Fig. 6).

Task 1: the participants had to complete a complex visual and textual fact-finding task involving multiple e-lectures using both interfaces (*Fact-finding* in Fig. 6).

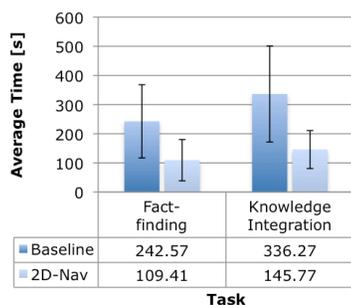


Figure 6. Average Time per Task for Inter-Video Navigation

Task 2: the participants were asked to complete a knowledge integration task for a given topic covered in multiple e-lectures (*Knowledge Integration* in Fig. 6). To exclude any learning effects, we used a between-subject design for the second task.

In both tasks, the participants were significantly faster ($p < 0.001$) using our video browser as shown in Figure 6. These results confirm that our user interface supports the user's orientation when navigating across multiple e-lectures. Moreover, statements in the interviews showed that the two dimensional browsing metaphor fosters the learners' awareness of interrelated e-lectures. The participants committed about 65% less usability errors using our video browser than using the baseline player (significant with $p < 0.001$).

Conclusion

The user interface concept presented in this paper is a first step towards supporting users in getting an overview of and navigating between multiple, semantically related videos in mobile scenarios. Our evaluation results show that the spatial concept supports (1) the user's orientation, (2) awareness of interrelations and (3) enables users to complete complex tasks significantly faster while committing significantly less usability errors than using a state of the art mobile video browser. Both, the grid-based layout and the vertical navigation were perceived as key concepts, improving the browser's usability. In order to further validate these results, we plan to conduct a long-term field study of our video browser.

Another promising way to deal with the limiting form factors of mobile devices is the extension of the user interface *beyond* the device. Further screen real estate

may be for instance available through interactive surfaces in urban pervasive environments or dynamically created using pico projectors in arbitrary mobile settings. Hence, as additional future work, we will investigate how we can leverage the unique affordances of these new technologies to further enlarge our spatial navigation metaphor.

Acknowledgements

This work was funded by the German Research Foundation (DFG-GK-1223).

References

- [1] Hare, J. S., Sinclair, P. A. S., Lewis, P. H., Martinez, K., Enser, P. G. and Sandom, C. J. Bridging the semantic gap in multimedia information retrieval: Top-down and bottom-up approaches. In *Proc. ESWC 2006*, Springer (2006).
- [2] Holland, J.D. and Hutchins, E. Opportunities and Challenges for Augmented Environments: A Distributed Cognition Perspective. In *Designing User Friendly Augmented Work Environments*, Springer (2009), 237-259.
- [3] Hürst, W., Welte, M., and Jung, S. An evaluation of the mobile usage of e-lecture podcasts. In *Proc. Mobility 2007*, ACM Press (2007), 16-23.
- [4] Hürst, W. and Meier, K. Interfaces for timeline-based mobile video browsing. In *Proc. MULTIMEDIA 2008*, ACM Press (2008), 469-478.
- [5] iTunes U. <http://www.apple.com/education/mobile-learning/>.
- [6] Kamvar M., Chiu P., Wilcox, L., Casi S. and Lertsithichai S. Minimedia surfer: browsing video segments on small displays. In *Proc. CHI 2004*, ACM Press (2004), 1371-1374.
- [7] MIT OpenCourseWare. <http://ocw.mit.edu/>.