

Wipe'n'Watch: Spatial Interaction Techniques for Interrelated Video Collections on Mobile Devices

Jochen Huber, Jürgen Steimle, Roman Lissermann,
Simon Olberding, Max Mühlhäuser
Technische Universität Darmstadt
Darmstadt, Germany

{jhuber,steimle,lissermann,olberding,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 *Wipe'n'Watch*, a novel user interface for the mobile navigation of large video collections comprising two 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.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces

General Terms

Design, Human Factors

Keywords

Multimedia, mobile devices, video browsing, interaction techniques, user interface, digital libraries

1. 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 billions of video streams e.g. through the iTunes store almost anytime and anywhere. Moreover, such devices typically have recording capabilities, therefore allowing us to record and share video data virtually 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) PocketDRAGON [6], which supports fine-grained in-scene navigation by direct manipulation on mobile devices.

We focus on video use at work, e.g. for learning on the job, not on entertaining videos watched during leisure time. Besides watching individual 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. The relationships are crucial for contrasting and integrating knowledge contained in related videos, therefore improving working efficiently. 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. In the following, we exemplify the shortcomings for a concrete application scenario: recordings of talks and lectures (so-called *e-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*). A video of the lecturer (*video stream*) is not necessarily presented due to its low information content [7]. The streams can be semantically segmented using the slides as key frames, each representing a semantic unit. The ubiquitous availability of multimedia learning material through services like iTunes U [8] or OpenCourseWare [9] has paved the way for groundbreaking changes in mobile learning. A recent study [5] 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. Furthermore, several topically related lectures can be used to gain deeper insight into a specific problem domain from a slightly different point of view. This practice is possible nowadays due to the vast amount of e-lectures available online from various universities.

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, (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, without

being aware of the interrelationships, having no overview over the actual lecture and supporting the navigation between e-lectures, it is impossible for learners to complete this task in a reasonable amount of time in a mobile setting.

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 *Wipe'n'Watch*, 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. 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.

2. INTERFACE 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 (see Fig. 1).

The horizontal dimension is used to browse within a semantically segmented video. The vertical dimension is used to navigate between topically related videos. The complex information space is hence mapped spatially onto the interaction space, fostering a user's overview.

2.1 Horizontal Navigation: Within a Video

The efficient navigation within an individual video and getting an overview on the video is crucial for knowledge work. For instance, knowledge workers must be able to easily find and access specific parts when reviewing contents, as well as to grasp the context of a particular topic in the scope of the video. These aspects require (1) getting *detailed information* on the current topic, (2) easy *navigation to related information* in the context of the current topic (e.g. preceding/following topics) and (3) an *efficient overview* on the entire video with quick access to any of the contents.

Since in practice these three activities are highly interrelated, we offer integrated support in one single interface. Instead of the timeline-based navigation of typical video

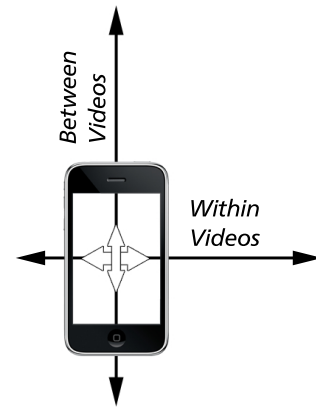


Figure 1: Spatial interaction concept

browsers, we utilize the key frames as basic navigation objects. These are advantageous for two reasons: First a key frame encapsulates coherent semantic content and second, it provides a good visual cue on its contents.

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: current topic and overview. The upper part shows the current topic in detail. Users can navigate through the key frames by simply wiping horizontally over the upper part of the user interface. Overview navigation within the entire video is supported in the lower part of the interface. This shows an overview with thumbnails of all key frames in a grid layout. The currently active key frame is highlighted. A key frame can be selected by tapping onto its thumbnail. Moreover, key frames can be skimmed very quickly by sliding the finger over the grid.

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 wiping horizontally.



Figure 2: Horizontal navigation within a video.

2.2 Vertical Navigation: Between Videos

Our concept for the navigation between topically related videos is based upon hyperlinks. These hyperlinks exist between semantic segments of the video (e.g. key frames). It is out of the scope of this paper how these links are created, since we focus on the navigation concept. Hyperlinks could be created automatically through multimedia information retrieval [2]. Furthermore, the user interface could be enhanced to allow users to manually create (and share) links between slides.

Our navigation support aims at providing an intuitive interaction technique, which allows users to follow hyperlinks and navigate easily within the navigation history. The major challenge hereby is to prevent users from getting lost in too much information presented on a small screen. Lost in Hypertext [1] is a well-known phenomenon, which may occur particularly in this situation. Due to this, we apply a spatial navigation concept: 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 wipes downwards, the interface is being scrolled downwards, revealing related videos as shown at the bottom of Figure 3a.



Figure 3: a) Vertical navigation between videos, b) Visualized browsing history

To provide an overview over the available related videos, they are aligned horizontally. 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. 3a, here: a news broadcast). 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 wiping vertically up and down respectively. Alternatively, to avoid repetitive wiping and to gain an overview on the browsing history, a visualization thereof can also be used for the vertical navigation as shown in Figure 3b. It is displayed as an image on top of the current video and can be navigated by moving the finger vertically across the images.

3. EVALUATION

We have implemented Wipe'n'Watch as part of a video browser for the Apple iPhone. It has been evaluated in a controlled experiment with 44 participants (30 male, 14 female) with different scientific backgrounds. Each single-user session lasted about 2 hours. The overall goal was to evaluate the effectiveness, efficiency, learnability and attractiveness [3] of the video browser, as well as user satisfaction.

The experiment was subdivided into two parts (within-subject). The first 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 usability and user experience goals. As data, we utilized recordings of lectures of each about 90 minutes length and news broadcasts. Prior to the experiment, we topically segmented the videos and manually created the interrelationships. The tasks of the participants comprised simple fact-finding tasks, as well as advanced knowledge integration tasks (see the following subsections). Both, time required to complete the tasks and usability errors were measured. For each task, a different set of videos was utilized to exclude any learning effects. The sessions were video-recorded and semi-structured interviews were conducted.

3.1 Intra-Video Navigation

The participants were presented three different user interfaces: (1) a slightly enhanced standard iPhone media player as baseline (Baseline in Fig. 4), which provided additional buttons to switch back and forth between key frames, (2) a player, which instead allowed users to skim through the key frames by wiping horizontally in landscape (Wipe only in Fig. 4) and (3) Wipe'n'Watch as shown in Fig. 2 including the overview grid, but without the possibility of inter-video navigation (W'n'W in Fig. 4). We introduced the wiping-only player to assess the particular influence of the horizontal wiping concept and to contrast it with the overview grid.

The participants were asked to complete three different fact-finding tasks with each user interface. The tasks required visual orientation within a video (task 1 and 3), as well as textual orientation (task 2), since the orientation and therefore a valid mental concept is crucial to quickly retrieve a desired part of a video. *Task 1:* the participants had to search an video for a given key frame without prior knowledge of the lecture (Visual 1 in Fig. 4). *Task 2:* the participants were asked to find a certain topic in the last third of the video (Textual in Fig. 4). *Task 3:* the participants had to navigate to the key frame following the one found in the first task (Visual 2 in Fig. 4).

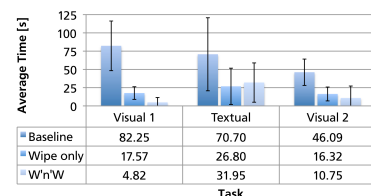


Figure 4: Average times for intra-video navigation

Figure 4 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 wiping-only browser or Wipe'n'Watch than using the baseline player. The participants also committed about 60% less usability errors (significant with $p < 0.01$).

Comparing Wipe'n'Watch with the wiping-only browser, we found that the participants were significantly faster using Wipe'n'Watch for task 1 ($p < 0.001$) and task 3 ($p < 0.05$). In task 2, the difference was not significant. This is in-line with qualitative findings from the semi-structured interviews. The participants stated that Wipe'n'Watch supports their visual orientation and navigation (as in task 1 and 3), whereas they prefer to skim through the key frames by wiping horizontally when they have no visual clues (as in task 2). Both, the wiping-only browser and Wipe'n'Watch were perceived as far more attractive (with an average score of 5 and 6 respectively) than the standard iPhone player (with a score of 2.5 on a 7-point Likert scale).

3.2 Inter-Video Navigation

The participants were presented two different user interfaces. First, a further enhanced standard iPhone media player (Baseline in Fig. 5), which allows switching forth and back between key frames, as well as browsing related videos using textual hyperlinks displayed on the key frames. Second, the participants were asked to utilize Wipe'n'Watch with both, horizontal and vertical navigation capabilities (W'n'W in Fig. 5).

The participants had to fulfill the following tasks. *Task 1*: the participants were asked to complete a complex visual and textual fact-finding task involving multiple videos using both interfaces (Fact-finding in Fig. 5). *Task 2*: the participants had to complete a knowledge integration task for a given topic covered in multiple videos (Knowledge Integration in Fig. 5). To exclude any learning effects, we used a between-subject design for the second task.

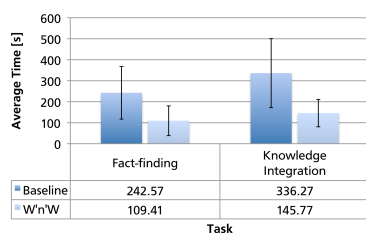


Figure 5: Average times for inter-video navigation

In both tasks, the participants were significantly faster ($p < 0.001$) using Wipe'n'Watch as shown in Figure 5. These results confirm that Wipe'n'Watch supports the user's orientation when navigating across multiple videos. Moreover, statements in the interviews showed that the two dimensional browsing metaphor fosters the users' awareness of interrelated videos. The participants committed about 65% less usability errors using Wipe'n'Watch than using the baseline player (significant with $p < 0.001$). Finally, Wipe'n'Watch was perceived as far more attractive with an average score of 6 than the baseline player with an average score of 3.5 on a 7-point Likert scale.

In the interviews, the participants commented on the spatial concept of Wipe'n'Watch as "clearly laid out" and they remarked that the vertical alignment of the related videos intensifies the visual relationship between the videos. This lets us draw the conclusion that the participants are more engaged in their working process using Wipe'n'Watch. Moreover, it supports them when deducing a mental concept of the videos.

4. SUMMARY

In this paper, we contribute Wipe'n'Watch. To the best of our knowledge, it is the first user interface concept for browsing videos on mobile devices that efficiently supports both navigating within single videos and between topically related videos. We have successfully shown how to cope with the limiting device characteristics by employing an efficient spatial navigation metaphor, which maps to the users' mental concepts.

The evaluation in a controlled experiment with 44 participants shows that our video browser significantly improves the working process by (1) supporting the user's orientation, (2) fostering awareness of interrelations and (3) enabling 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 horizontal and the vertical navigation were perceived as key concepts, improving the browser's attractiveness and usability, while the horizontal navigation was perceived as rather task-oriented and therefore improving the overall usability. In order to further validate these results, we plan to conduct a long-term field study of our video browser.

5. REFERENCES

- [1] D. M. Edwards and L. Hardman. *Lost in hyperspace: cognitive mapping and navigation in a hypertext environment*, pages 90–105. Hypertext: theory into practice. Intellect Books, Exeter, UK, 1999.
- [2] J. S. Hare, P. A. S. Sinclair, P. H. Lewis, K. Martinez, P. G. Enser, and C. J. Sandom. Bridging the semantic gap in multimedia information retrieval: Top-down and bottom-up approaches. In *Proc. ESWC '06*, 2006.
- [3] M. Hassenzahl, A. Platz, M. Burmester, and K. Lehner. Hedonic and ergonomic quality aspects determine a software's appeal. In *Proc. CHI '00*, pages 201–208, New York, NY, USA, 2000. ACM.
- [4] W. Hürst and K. Meier. Interfaces for timeline-based mobile video browsing. In *Proc. MM '08*, pages 469–478, New York, NY, USA, 2008. ACM.
- [5] W. Hürst, M. Welte, and S. Jung. An evaluation of the mobile usage of e-lecture podcasts. In *Proc. Mobility '07*, pages 16–23, New York, NY, USA, 2007. ACM.
- [6] T. Karrer, M. Wittenhagen, and J. Borchers. Pocketdragon: a direct manipulation video navigation interface for mobile devices. In *Proc. MobileHCI '09*, pages 1–3, New York, NY, USA, 2009. ACM.
- [7] T. Lauer, R. Müller, and S. Trahasch. Learning with lecture recordings: Key issues for end-users. In *Proc. ICALT '04*, pages 741–743. IEEE, 2004.
- [8] iTunes U. <http://www.apple.com/education/mobile-learning/>.
- [9] MIT OpenCourseWare. <http://ocw.mit.edu>.