VIDEO NAVIGATION ON TABLETS WITH MULTI-TOUCH GESTURES

Klaus Schoeffmann, Kevin Chromik, Laszlo Boeszoermenyi

Klagenfurt University, Universitaetsstr. 65-67, 9020 Klagenfurt, Austria
ks@itec.aau.at, kchromik@edu.aau.at,laszlo@itec.aau.at

ABSTRACT

We describe a new interaction method for video navigation on touch-enabled tablet devices, which is based on previous research results and uses context-sensitive swipe gestures. We evaluate our method in a user study with known-item-search tasks in direct comparison to seeker-bar navigation that is commonly used for navigation with video players on tablets and smartphones. Our evaluation results show that users prefer the swipe-based navigation feature over a seeker-bar in terms of convenience and that users can achieve better search performance with this new way of video navigation.

***Keywords*-video browsing, video navigation, mobile**

I. INTRODUCTION

Since their introduction in 2010, tablet PCs quickly became the ubiquitous personal computing devices used for many daily tasks, such as communication, web surfing, and photo management. According to a recent report [1], the global shipments of tablets per quarter grew faster than those of notebook computers and quickly exceeded the quarterly shipments of desktop computers. The same report shows an almost uniform distribution of the three classes (desktop, notebook, tablet) in terms of shipments per quarter in 2013.

Tablets have changed the way we interact with media, especially with photos. Instead of using a lot of buttons for these features, we simply wipe through our photos, pinch to zoom in/out of collections or into single picture details, drag to change the focus of a zoom operation and use multi-touch gestures to rotate images. Some applications even provide gyroscope-based navigation through photos or navigation by hand/finger gestures through optical tracking and a few other tools even use 3D interfaces for browsing [2].

However, interestingly the navigation within videos on tablet computers remained almost the same as on desktop computers. As shown in Figure 2 the same desktop-like interaction means are used: a seeker-bar and typical VCR-like control features. We argue that this is an inconsistent interaction concept and yields to confusion because tablet users are accustomed to wipe gestures for navigating through visual media. Therefore, we propose a video navigation method that uses context-sensitive wipe gestures to navigate forward and backward in videos. This general navigation method builds on existing work by combining the video navigation method proposed by Huber et al. [3] for e-learning videos (videos of lectures, including slide thumbnails) with the idea of the ZoomSlider, as proposed by Hürst et al. [4] for stylus-based mobile devices.

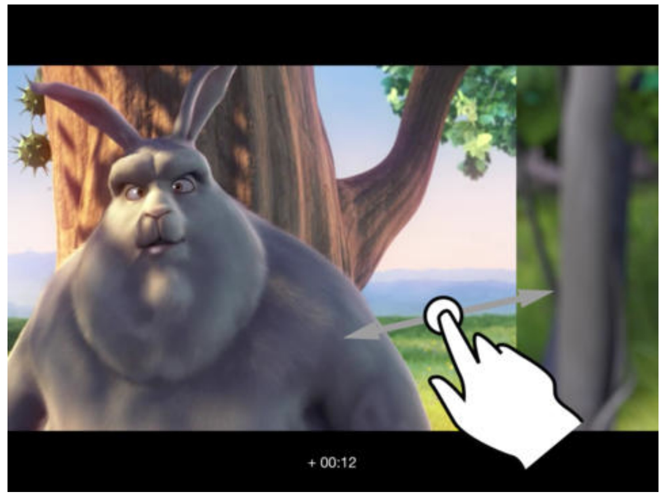

Fig. 1. Intra-video navigation: wipe left or right in order to temporally jump forward or backward. The vertical position of the wipe gesture defines the temporal step size of the navigation.

We want to know whether users prefer the wipe interaction over the seeker-bar interaction for navigation through videos and how well they perform for search tasks in comparison to target search with a typical, common video player. For that purpose we have developed a prototype application and tested it in a user experiment with 24 participants with known-item-search tasks against the default video player application on an Apple iPad. Our results show that users like the simple interaction concept and can efficiently use the provided navigation features, achieve a slightly better target search time with it (though not significantly better), and prefer the proposed player over the default video player.

II. RELATED WORK

Only a few papers address video navigation on mobile devices. Hürst et al. introduced the *Mobile ZoomSlider* interface [4] for – originally stylus-based – mobile handheld

To appear in *Workshop Proceedings of ICME 2014, July 14-18, 2014, Chengdu, China*

Fig. 2. Default controls of a mobile video player as used on the Apple iPad. In the lower part of the screen we can see fast-forward/reverse and play as well as volume control (which, however, can be controlled by other means, e.g., hard buttons, as well). In the upper part of the screen we can see the seeker-bar to be used for navigation by scrubbing. All controls are usually hidden during playback and need to be enabled by a tap gesture.

devices, which is an improved seeker-bar to allow both fine-grain and coarse-grain navigation in the video. The main idea is to use a virtual seeker-bar on any position on the screen and use the vertical click position as a parameter for navigation granularity. A drag to left or right direction results in backward or forward navigation and the vertical position of the drag operation defines how fast the navigation is performed. The same concept has been extended later to general mobile devices like PDAs and smartphones[5]. We borrow the idea of this concept and use it together with wipe gestures in videos to jump forward and backward according to a virtual scale that is coarse-grained at the top and fine-grained at the bottom (see Figure 1). However, in difference to the original navigation interaction we also move the video frame together with the swipe interaction to immediately show to future or past frames – and to allow cancelling the navigation operation as soon as necessary.

Karrer et al. propose the PocketDRAGON interface [6] to avoid occlusion of video content while navigating inside the video. Instead of showing an overlaid seeker-bar, as used with typical video players on mobile touch devices, they propose to directly manipulate objects in the scene for navigation, as also suggested by Dragicevic [7] earlier. For that purpose, motion tracking is performed on the video and object motion is used as basis for the dragging operation along a motion trajectory. In addition to this object-based navigation mode, which rather improves navigation accuracy and typically does not allow for quick navigation over longer segments in the video, they also support two finger gestures, where a horizontal swipe allows jumping to the previous/next scene. As no evaluation has been performed by the authors, it remains unclear how well this kind of navigation supports interactive search in videos.

To improve navigation in e-learning rather than entertaining videos, Huber et al. propose the *Wipe'n'Watch* interface for more convenient navigation in learning videos on mobile touch devices. Instead of a timeline navigation with a seeker-bar they propose to use a wipe gesture to navigate between keyframes, which in the case of e-lectures are the positions new slides appear. Their interface – which operates in portrait mode – is subdivided into two parts: (i) the upper part shows the actual video content and (ii) the lower part shows an overview of all keyframes, i.e., slides, that act as direct access points. Their work also targets inter-video navigation, similarly to the idea of the *RotorBrowser* proposed by De Rooij et al. for desktop use[8]. Thus, vertical wiping allows to jump between semantically similar segments among videos; e.g., topically related segments. The availability of such related segments is indicated with an arrow in the upper right corner of the interface.

III. INTERACTION CONCEPT

We propose an interaction concept that actually combines the idea of the *Mobile ZoomBrowser* with the idea of the *Wipe'n'Watch* interface and makes it usable for any kind of video genre. Instead of using a seeker-bar for timeline navigation, we allow backward and forward navigation by a horizontal wipe gesture. Thus, when the user touches the screen and gradually drags the finger to left direction (*wipe left*), the player will gradually move the current frame out of the screen (to the left) and at the same time gradually move a future frame into the screen from the right (Figure 3). This behaviour is directly bound to the movement of the finger, such that both (the current and future) frames are moved exactly as far as the user moves the finger on the screen (i.e., the user has the feeling the video frame sticks on his or her finger). This kind of navigation is consistent with the navigation in other media on mobile touch devices, such as navigation in photo albums, in presentation slides, and even in the surf history of mobile web browsers, which all use horizontal wipe gestures for forward and backward.

Our proposed method for video navigation does not use the velocity of the wipe gesture nor a rubber-band effect – as proposed by Huerst[5] – to define the step size of the navigational jump. However, in order to allow slow and fast navigation in both short and long videos, the *vertical position* of the horizontal wipe gesture is used as basis for the temporal granularity of the navigation. Instead of using a linear mapping of the vertical position to the jump size, we divide the overall height of the video content screen into five equidistant vertical areas (see Figure 4). This should allow for more convenient usage and help the user to roughly

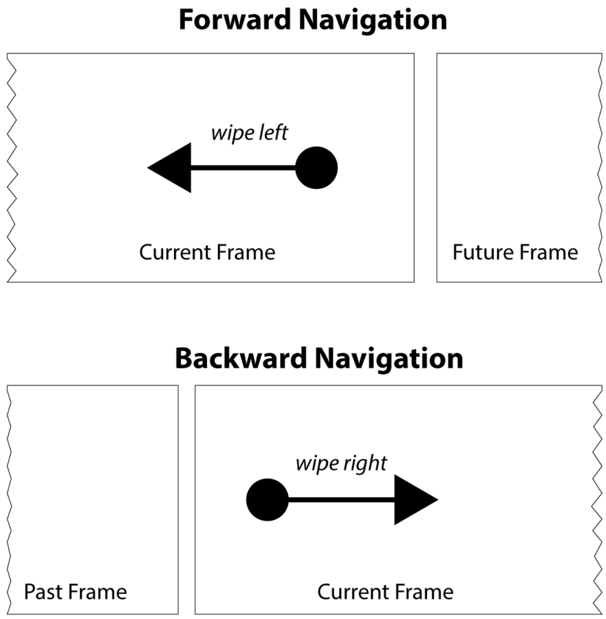

Fig. 3. Forward navigation (above): when the user drags the finger to the left, the current frame will move left out of the screen and a future frame will move into the screen from the right. Backward navigation (below): when the user drags to the right, the current frame will move right out of the screen and a past frame will move into the screen from the left.

estimate where to wipe for a specific step size. Wipe gestures performed in the lower area will result in backward (wipe right) or forward (wipe left) jumps by 0.5% of the entire video length. The jump size exponentially grows with each vertical area above.

In order to give a better understanding let us consider an example scenario where a user starts watching a 60 minutes video. If exactly at the time position 00:01:00 the user wants to jump forward and therefore performs a wipe gesture in left direction, he or she will jump by 18 seconds (0.5%) if the wipe gesture has been performed in the lower part of the screen (lowest of the five areas). However, if the left wipe is performed in the middle of the screen, the player will jump forward by 72 seconds (2%). Similarly, if the user performs a wipe to the right position in the top area of the screen, the player will jump backwards by 4.8 minutes (5%). To make it obvious how far the jump will be performed, our implementation displays the temporal offset of the jump during the wipe operation (see Figure 1).

It should be noted that a navigation action is not performed immediately as this could cause confusion for the user, because the temporal context is not preserved. Instead, the navigation is only performed if the user drags the finger far enough on the screen (e.g., at least 1/3 of the horizontal frame size). Moreover, a user can always cancel a naviga-

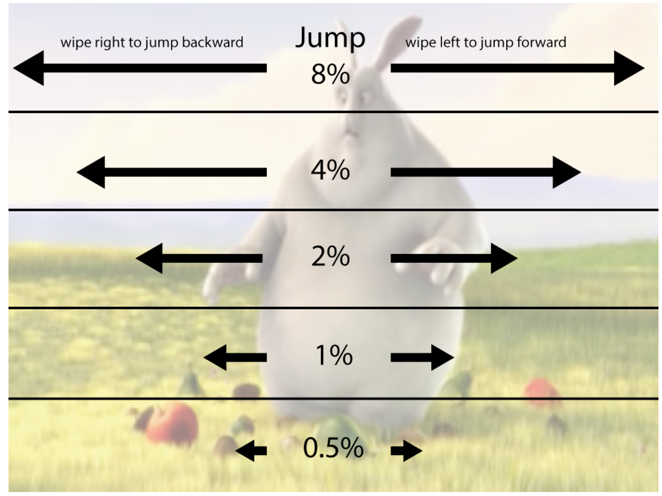

Fig. 4. The step size of the navigation within a video by a horizontal wipe gesture is defined by the vertical position, where the wipe gesture started. The overall height of the video screen is divided into five vertical regions that cause temporal jumps from 0.5% to 8% of the entire video length.

tion action by dragging back, such that the resulting wipe distance is below 1/3 of the frame width. This is a usual navigation behaviour on mobile touch devices and therefore should be immediately understood by the user.

Fig. 5. Inter-video navigation: vertical wipes are used to switch to different videos.

Similarly to the horizontal navigation, we use vertical wipe gestures for browsing through the available video files on the mobile device (in a ring manner). This should avoid additional interaction steps, like going back to the list of available videos in order to select one, and make the overall video browsing application more convenient Hence, a user can just wipe up or down to quickly circle through the available videos (see Figure 5). The previous playback/navigation position is preserved during this operation. So when the user visits the video again, it will continue at the position left off.

Fig. 6. The playback of the video can be controlled by a two-finger gesture.

Moreover, our video browser allows to pause and resume playback by a simple multi-tap gesture with two fingers (see Figure 6). Therefore, our video browsing application does not need any additional interaction controls that would limit the space for the video content or occlude it. The direct interaction by gestures also allow for immediate control and therefore facilitates quick navigation in contrast to the default video players on mobile devices, which use controls that are hidden to not limit or occlude the content and need to be brought to screen first by a tap gesture.

Finally, it should be noted that the interaction and navigation by gestures only has an additional advantage. People with limited vision should be able to easily use our interaction concept as it does not rely on small control widgets and textual information.

IV. USER EXPERIMENT

IV-A. Test Setup

To evaluate the efficiency of the proposed video navigation interface we performed a user study with 24 participants (two females) aged from 18 to 34, mean 25.63 years, s.d. 3.5. Out of the 24 users, 19 participants (79.16%) stated that they use a smartphone or tablet on a regular basis. One third of the users either wear glasses or contact lenses. Each user had to perform four known-item-search tasks in news video content, two of them with the default video player on an iPad and the other two with an app that implemented the interface concept described in Section III. We used a latin-square principle with random order to avoid memorisation effects. In order to not measure a single task performance but rather a general search performance in videos and to avoid any familiarisation effects, we used four video files (of about one hour duration) and defined four similarly located target segments for each video. Each user got a random selection of four tasks of different time position, with a uniform distribution over all users. As a device we used an iPad 3, equipped with 16 GB memory and WiFi, running iOS 6.1.3.

For each task the user has been presented with a 20 seconds video clip showing a segment of interest that had to be found as fast as possible. There was no time limit for a search but we allowed the users to abort a task if they could not solve it. A test session lasted for about 25 minutes in average, including a short introduction to the task the usage of the two video players. After the test each participant was requested to complete a questionnaire with SUS rating (System Usability Score) about the proposed player.

IV-B. Error Rate

As our test videos were rather long for interactive search tasks on a mobile device (all had a duration of about one hour), we did not force participants to finish a trial. Instead, we allowed a user to cancel a search task – for any reason whatsoever – if she could not find it. An analysis of the log files revealed that participants hard a hard time in finding the target scenes: 8 of 48 trials (16.7%) could not be found with the proposed player. However, the default video player with 10 unfound trials (20.8%) performed even worse in our study. Unfortunately, until now we could not find the reason for this rather high error rate.

IV-C. Navigation Behaviour

We wanted to learn how the participants of our study use the navigation possibilities provided by our implementation. Therefore, we logged every navigation interaction, i.e., the *step size* and *navigation direction*, performed by a user. We assumed that users would first start with coarse navigation (i.e., use the *jump by 8% of video duration* area in the top part of the screen) and navigate back and forth to narrow down the target segment and then switch to fine-grained forward search. Interestingly, most navigation actions were performed with the lowest granularity (see top part of Figure 7 and in Figure 8). However, when taking a look at the effective navigation distance, achieved through using specific step sizes, we can see that most navigation distance in the video had been performed with the mid-level step size of 4% of the video duration (bottom part of Figure 7). The values in this figure are based on the average video length of the study, which was almost exactly 60 minutes. As the figure shows a characteristic close to normal distribution, this is good evidence that the step sizes of the different vertical layers were configured with a good setting (at least for videos of 60 minutes duration) and the users did efficiently use our navigation feature with different levels of granularity.

The 24 participants of our study mainly used forward navigation. Overall we logged 2436 forward navigation

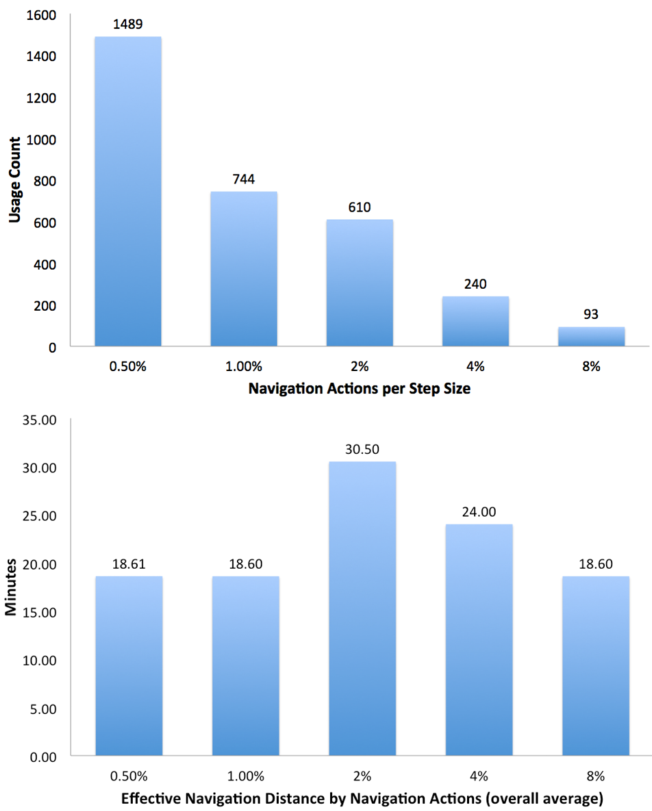

Fig. 7. Above: histogram of navigation actions per step size over all users and tasks. Below: Average navigation distance in a video (in minutes) per person achieved with the corresponding navigation step size.

actions (77%) and only 740 (23%) backward navigation actions. This behaviour is similar to the findings reported by a recent paper [9], where the authors investigated search behaviour with web-based video players. However, Figure 8 also shows that the chosen navigation granularity heavily varied over the different users. While participant 4 performed 313/44/1/0/1 navigation actions at granularity level 1/2/3/4/5 (where 1=0.5% and 5=8%), participant 6 achieved a very different distribution of 4/2/34/2/0. Both participants, however, were not able to find one of the two target scenes with the wipe-based video player.

The navigation diagram in Figure 9 also shows that users behaved were differently but stuck to their navigation behaviour. While user 11 navigated very fast forward and made only short jump backs, user 8 navigated very slowly and linearly. User 9 navigated forward only but made some changes of navigation speed (i.e., step size). Only user 10 showed the originally expected behaviour of quickly skimming through the video with coarse granularity, making larger jumps back and finally oscillated around the target position (also visible in Figure 8), finally found it but was much slower than user 11.

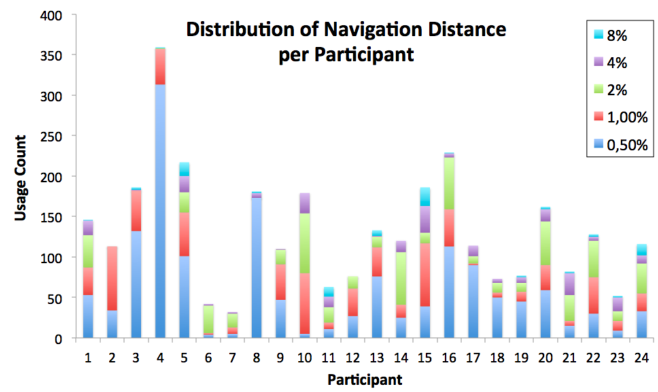

Fig. 8. Distribution of used navigation step size per participant

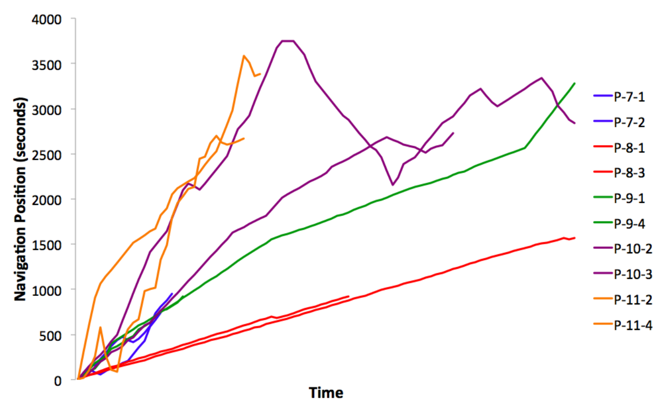

Fig. 9. Navigation diagram for both tasks of participants 7-11.

IV-D. Search Time

In terms of median search time the proposed wipe-based player (called *wipe player* in the following) performed slightly better overall, as also visible in Figure 10. It also showed lower variance, but a t-test showed no significant main effects due to too large variance within each interface ($t(22) = 1.171, p < 0.254$). However, interestingly enough, for the wipe player the variance in search time was very low for tasks with target positions in the first half of the video (see Figure 11). We can also see in the figure that the mean search time increases with the target position in the video. In contrast, the mean search time with the default video player rather tends to stay stable and shows similarly high variance over all target positions (though particularly high for the first task with the target position in the first quarter of the video). We speculate that this result was caused by too coarse navigation granularity of the seeker-bar of the default video player, such that the target segment had been overlooked.

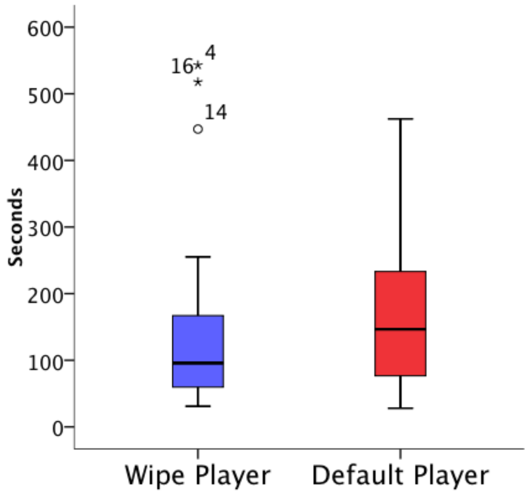

Fig. 10. Box plot of search time over all tasks and users

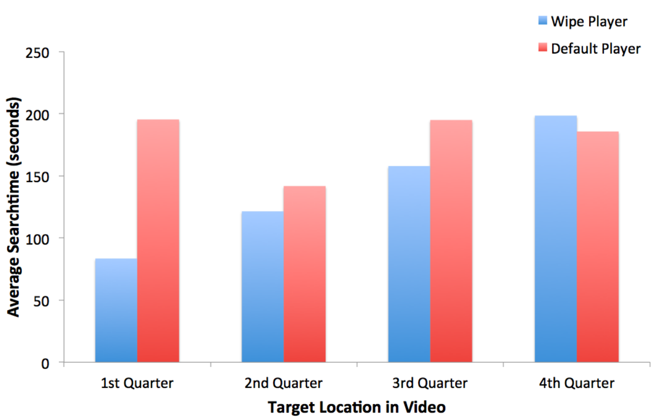

Fig. 11. Average search time per target location in the video.

IV-E. SUS Score

After the four search tasks each participant was asked to fill out a standardised *System Usability Scale* questionnaire as well as to assess which player gave more assistance in answering the search tasks. The evaluation of the SUS questionnaire resulted in an SUS score of 90. Overall the usability score – including the generally suggested correction of inverse rating due to control questions – had a very positive skew, with an mean average of 3.60 on a scale of 1 ('*strongly disagree*') to 5 ('*strongly agree*'). Particularly high ratings were given for the easiness of the wipe player (3.71) as well as the consistency (3.67) and the fast learnability (3.71). When asked about the preferred player, 21 out of 24 participants (87.5%) voted for the wipe player.

V. CONCLUSION

We have proposed a new interaction concept for videos on mobile devices using multi-touch screens. The provided navigation features enable users to wipe in videos from one scene to another by a simple touch gesture that is also used for content switching by default image viewers and presentation slide viewers on these devices. We have implemented a prototype of this video player and tested it in a user study with 24 users that performed known-item-search tasks in videos of one hour length. Our results show that the participants could effectively use the provided navigation features to find the desired target scenes. Moreover, for target scenes located in the first 30 minutes of the videos they could even perform faster search than with a common video player.

More than 87 percent of users in our study reported that the wipe player could better help them in finding the desired target segments. From a usability perspective they highly favoured the consistency, fast learnability, and easiness of the wipe player. We have also received very positive feedback from external users (not participating in the study) because of the fact that the player uses not the typical interaction controls but only gestures.

ACKNOWLEDGMENT

This work was funded by the Federal Ministry for Transport, Innovation and Technology (bmvit) and the Austrian Science Fund (FWF): TRP 273-N15 and the European Regional Development Fund and the Carinthian Economic Promotion Fund (KWF), supported by Lakeside Labs GmbH, Klagenfurt, Austria.

VI. REFERENCES

[1] M. Meeker and L. Wu, "2013 internet trends," May 2013. [Online]. Available: http://www.kpcb.com/insights/2013-internet-trends

[2] D. Ahlström, M. A. Hudelist, K. Schoeffmann, and G. Schaefer, "A user study on image browsing on touchscreens," in *Proceedings of the 20th ACM International Conference on Multimedia*, ser. MM '12. New York, NY, USA: ACM, 2012, pp. 925–928. [Online]. Available: http://doi.acm.org/10.1145/2393347.2396348

[3] J. Huber, J. Steimle, R. Lissermann, S. Olberding, and M. Mühlhäuser, "Wipe'n'watch: spatial interaction techniques for interrelated video collections on mobile devices," in *Proceedings of the 24th BCS Interaction Specialist Group Conference*, ser. BCS '10. Swinton, UK, UK: British Computer Society, 2010, pp. 423–427. [Online]. Available: http://dl.acm.org/citation.cfm?id=2146303.2146367

[4] W. Hürst, G. Götz, and M. Welte, "Interactive video browsing on mobile devices," in *Proceedings of the 15th international conference on Multimedia*, ser. MULTIMEDIA '07. New York, NY, USA: ACM, 2007, pp. 247–256. [Online]. Available: http://doi.acm.org/10.1145/1291233.1291284

[5] W. Hürst and K. Meier, "Interfaces for timeline-based mobile video browsing," in *Proceedings of the 16th ACM*

international conference on Multimedia. ACM, 2008, pp. 469–478.

[6] T. Karrer, M. Wittenhagen, and J. Borchers, "Pocketdragon: a direct manipulation video navigation interface for mobile devices," in *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, ser. MobileHCI '09. New York, NY, USA: ACM, 2009, pp. 47:1–47:3. [Online]. Available: http://doi.acm.org/10.1145/1613858.1613917

[7] P. Dragicevic, G. Ramos, J. Bibliowitcz, D. Nowrouzezahrai, R. Balakrishnan, and K. Singh, "Video browsing by direct manipulation," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '08. New York, NY, USA: ACM, 2008, pp. 237–246. [Online]. Available: http://doi.acm.org/10.1145/1357054.1357096

[8] O. de Rooij, C. G. M. Snoek, and M. Worring, "Mediamill: semantic video search using the rotorbrowser," in *Proceedings of the 6th ACM international conference on Image and video retrieval*. ACM Press, 2007, pp. 649–649. [Online]. Available: http://doi.acm.org/10.1145/1282280.1282376

[9] K. Schoeffmann and C. Claudiu, "An evaluation of interactive search with modern video players," in *Proceedings of the IEEE International Conference on Multimedia and Expo Workshops (ICMEW'13)*, 2013, p. 4.