

# An Evaluation of Video Browsing on Tablets with the ThumbBrowser

Marco A. Hudelist and Klaus Schoeffmann

Klagenfurt University, Universitätsstrasse 65-67, A-9020 Klagenfurt, Austria,  
{marco,ks}@itec.aau.at

**Abstract.** We present an extension and evaluation of a novel interaction concept for video browsing on tablets. It can be argued that the best user experience for watching video on tablets can be achieved when the device is held in landscape orientation. Most mobile video players ignore this fact and make the interaction unnecessarily hard when the tablet is held with both hands. Naturally, in this hand posture only the thumbs are available for interaction. Our *ThumbBrowser*-interface takes this into account and combines it in its latest iteration with content analysis information as well as two different interaction methods. The interface was already introduced in a basic form in earlier work. In this paper we report on extensions that we applied and show first evaluation results in comparison to standard video players. We are able to show that our video browser is superior in terms of search accuracy and user satisfaction.

**Keywords:** video browsing; mobile computing; human computer interaction

## 1 Introduction

Using tablets in landscape orientation when watching videos can be quite cumbersome. User interfaces of mobile video players usually ignore the typical position of the hands at the sides of the device. Although mobile touchscreen devices offer manifold interaction options, video players still look and work very similar to their desktop counterparts, as can be seen in Fig. 5. Holding a device in landscape orientation for longer periods of time is usually more convenient, in particular for large tablets, e.g., 12-inch devices. As a result, interaction with the players' interface becomes unnecessarily hard. Using playback controls that are positioned in the middle of the screen like seeker bars, play/pause/fast forward/fast rewind buttons or similar, can cause strains of the hands and the fingers. This problem is ignored even though there is an increasing amount of video content that is watched with mobile video players. Furthermore, embedded camera systems and mobile content delivery have greatly improved and more video content – for professional and entertainment purposes – is consumed than ever before.

To solve this problem an ergonomic video player interface should be designed that is tailored for the special needs in mobile multimedia interaction. As all

eight fingers are occupied when holding the device in landscape orientation, the thumbs become especially important. In contrast to the fingers they are free to move at the left and right sides of the device. Therefore, UI controls need to be positioned at these areas on the screen, so that all functions can be easily reached with the thumbs. The idea of such a UI layout is in fact far from new. On-screen keyboards offer a special mode for such use cases already for a long time. The keyboard is split up into two pieces and placed at the left and right side of the screen. In Fig. 2 examples of default iOS and Windows keyboards can be seen. Nevertheless, this idea is completely ignored by user interfaces of current mobile video players.



Fig. 1: Interface of the ThumbBrowser.

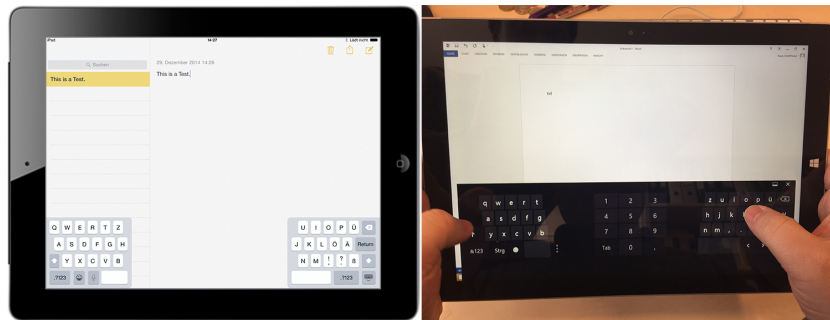


Fig. 2: Examples of split keyboards on iOS and Windows.

In [9] we already proposed to use the thumbs for mobile video interaction. However, at that time we were only able to present a first, basic version of our ThumbBrowser. As we took part in the demo paper track we also were not able to provide extensive evaluation results and therefore had to leave the readers unclear about a definite benefit. Therefore, in this work we build on this basis and present extensions to our interface as well as results of a comparative user evaluation with a state-of-the-art mobile video player. The evaluation configuration is inspired by the work of Schoeffmann and Burgstaller [16], as they performed a similar experiment with a video player interface tailored for smartphone screens in portrait orientation. They utilize the idea of a scrubbing wheel, as used on Apple iPods to control music, for navigating in videos. Moreover, we are comparing our results to theirs in the analysis of the evaluation results.

## 2 Related Work

Hürst and Darzentas [11] propose browsing videos on a tablet with a hierarchical storyboard interface. A videos' content is represented by a grid of thumbnails representing segments that can be tapped to be transferred to another, lower-level grid, representing the content of the specific segment. Hürst et al. [12] also show mobile browsing for timeline-based video browsing with focus on PDA-like devices that are operated with a stylus. For example, users are able to control seek-speed by varying the vertical position of the stylus. This approach is similar to the interaction paradigm later adopted by the default iOS video player. Furthermore, Hürst et al. [13] present a concept for video navigation on smartphones tailored for one-handed, i.e. thumb interaction.

Hudelist et al. [8] utilize the metaphor of a 3D filmstrip for browsing videos on tablet device, similar to the motivation of [1]. The content is represented by a floating filmstrip as it is used in analog film projectors and cameras. Each image on the strip represents a video segment that can be directly played in the strip visualization. Ganhör [5] shows ProPane, an interface for fast and very precise mobile browsing on smartphone-like devices. It enables users to control playback and seeking in a very precise manner, e.g. for video editing scenarios. Huber et al. [7] present Wipe'n'Watch, an interface for browsing interrelated video collections, similarly to the approach of De Rooij et al. [3] for desktop computers. Karrer et al. [15] propose an interface for mobile devices that utilizes direct manipulation of objects in a scene for navigation instead of traditional seeker bars as was shown in earlier work by Dragicevic et al. [4] for desktop PCs.

Moreover, Schoeffmann et al. [17] show a mobile video player that uses wipe gestures for controlling the seeking speed. Hudelist et al. [10] also show a video player for navigation in single videos on tablet devices that utilizes sub-shots and different levels of detail for browsing via keyframes. For this, the interface uses three synchronized filmstrips with which users can easily navigate in the content. A purely human-computational approach is shown by Hürst et al. [14] where users browse through videos by inspecting a large array of uniformly sampled keyframes. Zhang et al. [19] present a mobile interface for collaborative browsing

of two users in a single video with the abilities to share sketches. The position in the video is controlled via simple touch gestures. Similarly, Cobarzan et al. [2] proposes a system for collaborative browsing with multiple mobile clients using tablets and a single server that manages communication and query requests of the clients. Finally, a general overview of the field of novel video browsing interfaces is given by Schoeffmann et al. [18].

### 3 Interface

The extended interface that is presented in this paper is based on our earlier work on the ThumbBrowser (see Hudelist et al. [9]). We extended our original concept with functionality regarding content analysis and user interaction, before we performed a comparative user study to prove its usefulness to users.

The interface tries to avoid interfering with users' watching experience as much as possible. Therefore, in normal playback mode all of the UI controls are hidden. When users want to interact with the player, e.g., to change playback position, they simply have to put one of their thumbs on either side on the screen. Depending on which side is touched different controls become visible, as can be seen in Fig. 1.

On the right hand side a vertical seeker control is activated. It is inspired by the classical layout of a traditional seeker bar but the timeline is positioned vertically. As a result, users can easily navigate to every position inside a video with their right thumb as every part of the timeline is reachable. Furthermore, a preview window appears in relation to users vertical thumb position on the screen. This feature is similar to the magnifying glass functionality shown by latest iterations of video players used by YouTube and similar websites. Therefore, to jump to any position in the video users have to place their right thumb on the screen, drag it to the wanted position on the timeline and then lift it up again. This activates the navigation process and the playback position is changed accordingly. It is possible to avoid this navigational jump by dragging the thumb all way to the right out of the screen, instead of lifting it.

We further extended the original interface concept by automatically analyzing the current video in the background and determining the dominant color in five second steps. This is done by creating a simple color histogram based on the HSB color space. Color values are assigned to one of eight bins. The bins cover value ranges of different size and were defined by results of a preliminary test.

On the left side of the screen a radial menu can be activated. It offers options to play/pause the video, perform fast forwarding and fast rewinding.

Additionally, we added another timeline visualization mode to the interface, called the *filmstrip mode*.

The *filmstrip mode* provides a scroll-able list of keyframes at the right hand side of the screen, as can be seen in Fig. 4. The keyframes are uniformly sampled from the video in five second steps.

When users tap on one of the keyframes the video player adjusts accordingly. This feature is designed to help users refine their search in case of very long

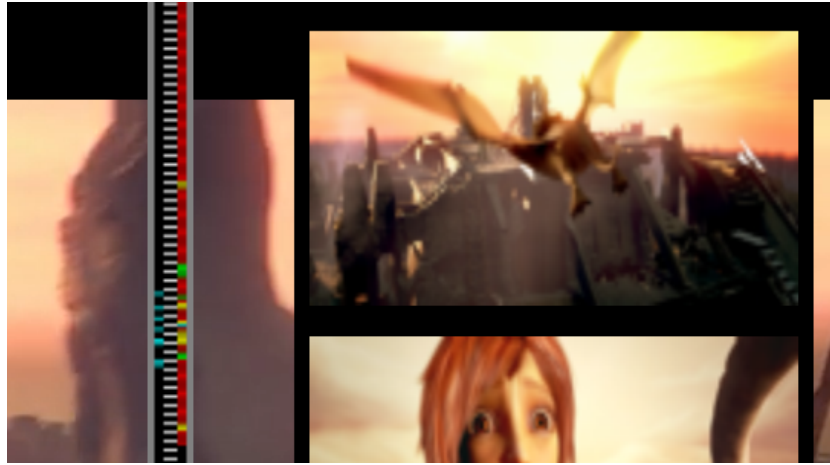


Fig. 3: Visualization of dominant color directly in the vertical timeline.

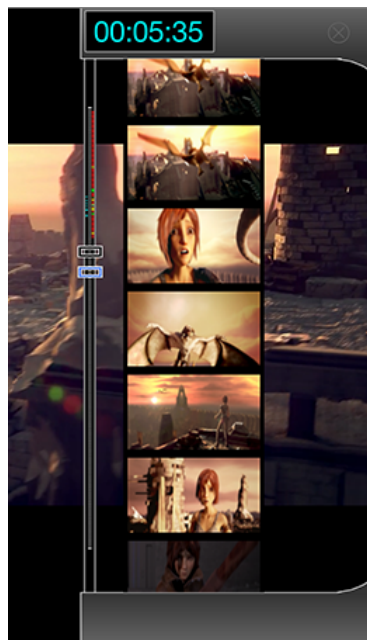


Fig. 4: Vertical seeker control after activation of the filmstrip mode.

videos. For example, they start a rather crude search with the seeker control and after some time notice a promising section of the video. As the seeker control is too sensitive to examine it in detail they are able to continue their browsing process by switching to the filmstrip mode.

## 4 Evaluation

In order to make the results of this study comparable we designed the evaluation to be very similar to the one used by Schoeffmann and Burgstaller [16]. We even used the same data set (we want to thank the authors for providing the whole data set with ground truth).

Participants of the user study had to search and mark all occurrences of pre-defined objects. Four videos had to be inspected and annotated. The first video was a documentary about gravity and planets in outer space with a duration of 35 minutes. In this video users had to find all scenes where images of the planet Earth were visible. The second video was an extended report about worldwide multi-cultural societies with a duration of 30 minutes. In this case users had to find scenes where glasses were shown. The third video was a documentary about cultivating fruits and vegetables with a duration of 25 minutes where study participants had to find all occurrences of bananas. Finally, video four was a report about *Gamification* with 40 minutes of length where all scenes with smartphones had to be marked by the participants.

Moreover, participants were told that they could spend as much time as they deemed appropriate to complete a task, but it was recommended not to spend more than up to seven minutes in each case. This was done in order to avoid putting too much stress on users by trying to find really every instance and thus spending unrealistic amounts of time.

To compare the performance of the ThumbBrowser to the performance of a default player we used the media player control of the iOS API, which offers a play/pause button and a seeker bar. Furthermore, two buttons were always available in both interfaces: one button to mark a scene and one button to finish the current task. As testing device we used an iPad Air (first generation) with iOS 9.3.

The order of interfaces and videos were alternated between each participant with the exception that two consecutive videos were always tested with the same interface, e.g., video one and two were processed with interface A, followed by video three and four processed with interface B. Furthermore, before a user could start the study, we asked them to provide us with their age, gender and smartphone/tablet experience level, e.g., beginner, advanced or professional. At the beginning of each task the system described the required objects textually. After completion of the first two tasks with the first interface a questionnaire was displayed where users had to give ratings about the interface according to the NASA Task-Load-Index (TLX) [6] and users had to rate them on a Likert-scale. The following questions were asked: (i) how mentally demanding the interaction was, (ii) how physically demanding the interaction was, (iii) how much they had

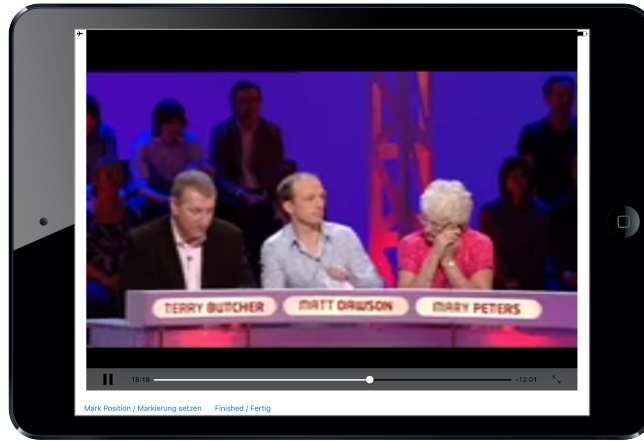


Fig. 5: Screenshot of the standard video player.

the feeling that the interface supported them in solving the task, (iv) how much fun it was to use the interface, (v) how frustrating it was to use the interface, and (vi) how easy to understand and easy to use the interface was.

#### 4.1 Experimental Results and Statistical Analysis

In total 26 participants took part in the study of which exactly the half were female. Average age of the participants was at 25.5 years. Moreover, eight indicated that they were smartphone/tablet beginners, 13 told us that they were advanced users and five selected that they were very experienced users.

A paired-samples t-test was used to determine whether there was a statistically significant mean difference between the search performance, e.g. how many scenes were found with the ThumbBrowser compared to the standard player. One outlier was detected that was more than 1.5 box-lengths from the edge of the box in a boxplot. Inspection of its value did not reveal it to be extreme and it was kept in the analysis. The difference scores for the search performance of the ThumbBrowser and standard player were normally distributed, as assessed by Shapiro-Wilk's test ( $p = 0.70$ ). Data are mean  $\pm$  standard deviation, unless otherwise stated. The percentage of found scenes was higher for the ThumbBrowser ( $66.19\% \pm 21.28\%$ ) than for the standard player ( $48.43\% \pm 18.68\%$ ). The test revealed a statistically significant difference between the two interfaces ( $t(51) = 4.53$ ,  $p < 0.0005$ ,  $d = 0.63$ ). Please see Fig. 6 for a visualization of the performance differences.

This result is encouraging as it shows that the ThumbBrowser can in fact improve users performance significantly. Next, we analyzed the differences on how much time participants spent for each task.

A paired-samples t-test was used to determine if there was a statistically significant mean difference between the search times of the ThumbBrowser compared to the standard player. Three outliers were detected that were more than

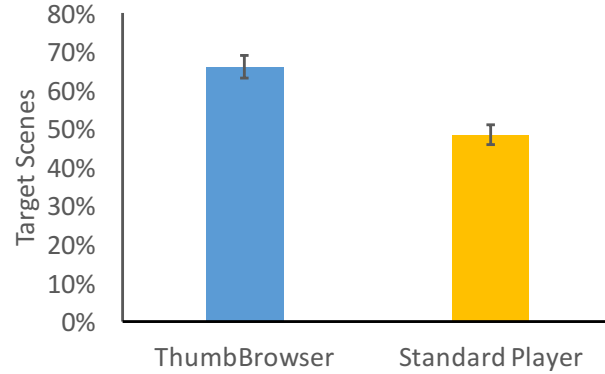


Fig. 6: Amount of retrieved target scenes (error bars:  $\pm$  s.e. of the mean).

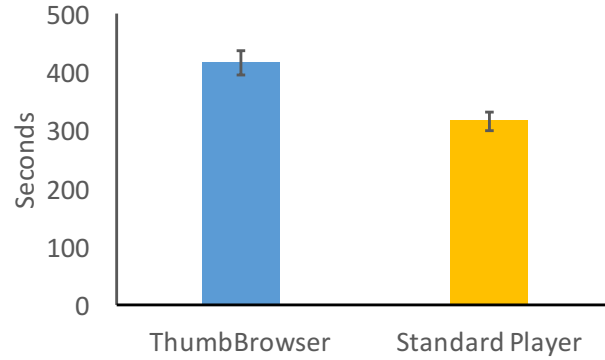


Fig. 7: Task solve time (error bars: 95% confidence interval).

1.5 box-lengths from the edge of the box in a boxplot. Inspection of their values did reveal that one of them was extreme and therefore was excluded from the analysis. The difference scores for the ThumbBrowsers' search times and the standard players' search times were normally distributed, as assessed by Shapiro-Wilk's test ( $p = 0.16$ ). Data are mean  $\pm$  standard deviation, unless otherwise stated. The search times were higher for the ThumbBrowser ( $417.1s \pm 217.4s$ ) than for the standard player ( $315.7s \pm 229.9s$ ). The test revealed a statistically significant difference between the two interfaces ( $t(50) = 2.529$ ,  $p < 0.05$ ,  $d = 0.35$ ). In Fig. 7 the difference in mean search times are visualized.

This result is interestingly similar to the results of Schoeffmann and Burgstaller [16] and could indicate that users were more comfortable with the ThumbBrowser than with the standard player. Therefore, they invested more time in the tasks and this could also contribute to the overall better search performance.

To determine if there are statistical significant differences between the answers given in the questionnaires a Wilcoxon signed-rank test was performed.



In case of mental demand the ThumbBrowser was significantly less demanding than the standard player ( $Z = 2.22$ ,  $p < 0.05$ ). Moreover, it was physically less demanding to use ( $Z = 2.373$ ,  $p < 0.05$ ), users had the feeling that it significantly supported them in their tasks more than the standard player ( $Z = -4.373$ ,  $p < 0.005$ ), it was significantly more fun to use ( $Z = -4.286$ ,  $p < 0.005$ ) and less frustrating ( $Z = 4.106$ ,  $p < 0.005$ ). In terms of usability both interfaces were equal ( $Z = 0.776$ ,  $p = 0.438$ ). In Fig. 8 the differences between the interfaces regarding the questionnaires are visualized.

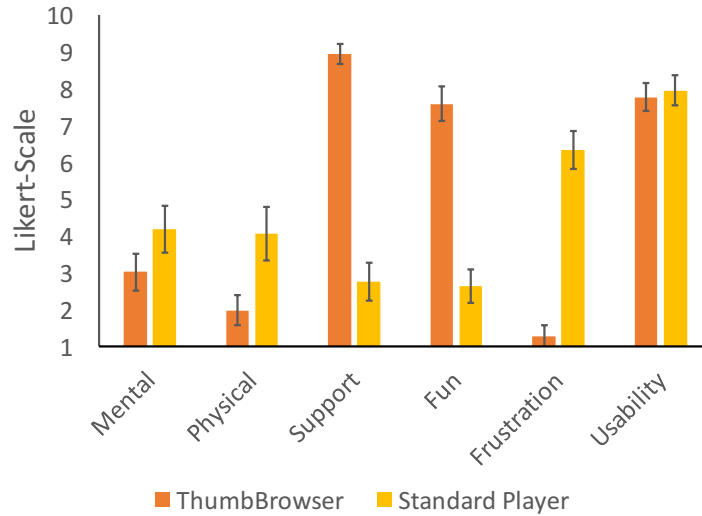


Fig. 8: Workload ratings according to NASA Task-Load-Index [6], with Likert-scale 1-10, for both interfaces (error bars:  $\pm$  s.e. of the mean). Lower is better for *Mental*, *Physical* and *Frustration*.

When we visually examined the navigation behaviors of the participants it also became clear that the search strategies between the two interfaces were slightly different (see Fig. 9). In case of the ThumbBrowser users were much more likely to perform their search unidirectional, e.g., starting from the beginning and searching to the end of the video. This was not the case for the standard player, as users were more likely to restart their search again and again from the beginning when they reached the end of the video. Therefore, we can agree with the findings of Schoeffmann and Burgstaller [16] who discovered the same behavior. Moreover, when we further compare their results to ours, we see that although the search performance as well as the average search duration were slightly higher in our study, the general trend is very similar. People spend significantly less time with the standard player and also find significantly less target scenes.

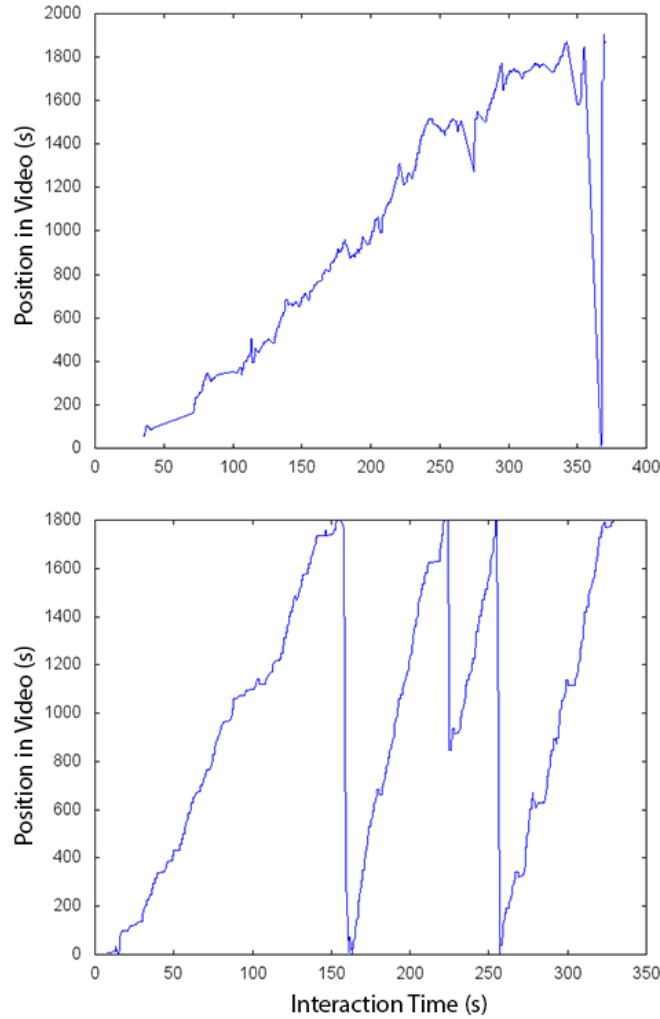


Fig. 9: Sample of the navigation behavior for one user in the documentary video. Users have to find scenes where planet earth is visible when using the Thumb-Browser (top) and the standard player (bottom).

## 4.2 User Feedback

Some additional comments that were given after completing the user study were that the ThumbBrowser *offered a much smoother seeking interaction* and that they would prefer *a different way to open and close the filmstrip UI*. Furthermore, they told us that *additional fast forwarding and fast rewinding speeds* would have also helped, as in its current iteration the ThumbBrowser only supports a fixed seeking speed of two times the normal playback speed. One participant who was

editing a lot of videos professionally told us that she would love to have the interface for her work.

## 5 Conclusions

In this paper we have presented and evaluated an extended version of the Thumb-Browser - a video browsing and video search tool for tablets that is optimized for landscape operation.

As playback controls are often hard to reach when holding the device with both hands at the sides, using interface controls explicitly designed for the thumbs provides a much better user experience. The interface provides a vertical seeker control similar to a timeline on the right hand side and a radial menu with additional playback functionality on the left hand side of the screen. Moreover, the extended version of the interface uses an easy to understand visualization of dominant colors across the video for faster navigation to scenes with salient color patterns and provides means to switch between coarse and detailed browsing modes.

We tested the interface in a user study with 26 participants where users had to mark scenes containing predefined objects. The results of our study show that the ThumbBrowser could outperform a traditional standard player by letting users find significantly more target scenes. Also, it is significantly less demanding in terms of mental and physical needs, it supports users better in solving the tasks, is more fun, less frustrating and as easy-to-use as a standard player.

## References

1. C. Beecks, T. Skopal, K. Schöffmann, and T. Seidl. Towards large-scale multimedia exploration. In *Proceedings of the 5th International Workshop on Ranking in Databases (DBRank 2011)*, pages 31–33, 2011.
2. C. Cobarzan, M. A. Hudelist, and M. Del Fabro. Content-based video browsing with collaborating mobile clients. In C. Gurrin, F. Hopfgartner, W. Hurst, H. Johansen, H. Lee, and N. O'Connor, editors, *MultiMedia Modeling*, volume 8326 of *Lecture Notes in Computer Science*, pages 402–406. Springer, 2014.
3. O. de Rooij, C. G. M. Snoek, and M. Worring. Mediamill: Semantic video search using the rotorbrowser. In *Proc. of the 6th ACM Int. Conf. on Image and Video Retrieval, CIVR '07*, pages 649–649, New York, NY, USA, 2007. ACM.
4. P. Dragicevic, G. Ramos, J. Bibliowicz, D. Nowrouzezahrai, R. Balakrishnan, and K. Singh. Video browsing by direct manipulation. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems, CHI '08*, pages 237–246, New York, NY, USA, 2008. ACM.
5. R. Ganhör. Propane: Fast and precise video browsing on mobile phones. In *Proc. of the 11th International Conference on Mobile and Ubiquitous Multimedia, MUM '12*, pages 20:1–20:8, New York, NY, USA, 2012. ACM.
6. S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati, editors, *Human Mental Workload*, volume 52 of *Advances in Psychology*, pages 139 – 183. North-Holland, 1988.

7. 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 *Proc. of the 24th BCS Interaction Specialist Group Conference*, BCS '10, pages 423–427, Swinton, UK, UK, 2010. British Computer Society.
8. M. A. Hudelist, K. Schoeffmann, and L. Boeszoermenyi. Mobile video browsing with a 3d filmstrip. In *Proc. of the 3rd ACM Conference on Int. Conf. on Multimedia Retrieval*, ICMR '13, pages 299–300, New York, NY, USA, 2013. ACM.
9. M. A. Hudelist, K. Schoeffmann, and L. Boeszoermenyi. Mobile video browsing with the thumbbrowser. In *Proceedings of the 21st ACM International Conference on Multimedia*, MM '13, pages 405–406, New York, NY, USA, 2013. ACM.
10. M. A. Hudelist, K. Schoeffmann, and Q. Xu. *Proc. of the 21st Int. Conf. on MultiMedia Modeling*, chapter Improving Interactive Known-Item Search in Video with the Keyframe Navigation Tree, pages 306–317. Springer, Cham, 2015.
11. W. Hürst and D. Darzentas. History: A hierarchical storyboard interface design for video browsing on mobile devices. In *Proc. of the 11th Int. Conf. on Mobile and Ubiquitous Multimedia*, MUM '12, pages 17:1–17:4, New York, NY, USA, 2012. ACM.
12. W. Hürst, G. Götz, and M. Welte. A new interface for video browsing on pdas. In *Proc. of the 9th Int. Conf. on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '07, pages 367–369, New York, NY, USA, 2007. ACM.
13. W. Hürst and P. Merkle. One-handed mobile video browsing. In *Proc. of the 1st Int. Conf. on Designing Interactive User Experiences for TV and Video*, UXTV '08, pages 169–178, New York, NY, USA, 2008. ACM.
14. W. Hürst, R. van de Werken, and M. Hoet. A storyboard-based interface for mobile video browsing. In X. He, S. Luo, D. Tao, C. Xu, J. Yang, and M. Hasan, editors, *MultiMedia Modeling*, volume 8936 of *Lecture Notes in Computer Science*, pages 261–265. Springer, 2015.
15. T. Karrer, M. Wittenhagen, and J. Borchers. Pocketdragon: A direct manipulation video navigation interface for mobile devices. In *Proc. of the 11th Int. Conf. on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '09, pages 47:1–47:3, New York, NY, USA, 2009. ACM.
16. K. Schoeffmann and L. Burgstaller. Scrubbing wheel: An interaction concept to improve video content navigation on devices with touchscreens. In *IEEE Int. Symposium on Multimedia (ISM)*, pages 351–356, Dec 2015.
17. K. Schoeffmann, K. Chromik, and L. Boeszoermenyi. Video navigation on tablets with multi-touch gestures. In *IEEE Int. Conf. on Multimedia and Expo Workshops (ICMEW)*, pages 1–6, July 2014.
18. K. Schoeffmann, M. A. Hudelist, and J. Huber. Video interaction tools: A survey of recent work. *ACM Comput. Surv.*, 48(1):14:1–14:34, Sept. 2015.
19. J.-K. Zhang, C.-X. Ma, Y.-J. Liu, Q.-F. Fu, and X.-L. Fu. Collaborative interaction for videos on mobile devices based on sketch gestures. *Journal of Computer Science and Technology*, 28(5):810–817, 2013.