SCRUBBING WHEEL: AN INTERACTION CONCEPT TO IMPROVE VIDEO CONTENT NAVIGATION ON DEVICES WITH TOUCHSCREENS

Klaus Schoeffmann and Lukas Burgstaller

Klagenfurt University
Institute of Information Technology
Universitätsstrasse 65-67, A-9020 Klagenfurt, Austria
ks@itec.aau.at, mail@voidstern.net

ABSTRACT

We propose a new interface that facilitates content navigation in videos on devices with touchscreen interaction. This interface allows both coarse-grained and fine-grained navigation in an intuitive way and enables better performance when used to locate specific scenes in videos. We implemented this interface on a 5.5-inch smartphone and tested it with 24 users. Our results show that for video navigation tasks the proposed interface significantly outperforms the seeker-bar interface, commonly used with video players on mobile devices. Moreover, we found that the interaction concept of the Scrubbing Wheel has a much lower perceived workload than the widely used seeker-bar, and is the preferred tool to locate scenes in videos for all tested users in our study.

Index Terms**—** Video Browsing and Search, Touch Interaction, Mobile

1. INTRODUCTION

Over the last few years we could observe a significant improvement of photo and video capturing capabilities of smartphones and tablets. State-of-the-art smartphones nowadays support image capturing with 16 MP, including sensor-based image stabilization, as well as 4K video recording at very high frame rates (e.g., 120 fps). At the same time they provide multi-touch interaction with high resolution screens, such as Quad HD and 577 pixels per inch.

However, despite the incredible improvements in record and display functionality, the video interaction software still lags behind. As illustrated in Figure 1, the standard interface of mobile video players still uses the very same interaction means as used by video players on desktop PCs for several decades already: buttons for *play*, *pause*, *jump forward/backward*, and a *seeker-bar for navigation* (i.e., random access) in video. Though on iOS systems the seeker-bar provides a *fine scrubbing* feature that allows to change the navigation resolution by dragging the virtual seeker-bar to different vertical position (see the textual hint below the seeker-bar at right top of Figure 1), navigation in video on smartphones is still very inconvenient and cumbersome. This is especially true when the smartphone is held in portrait orientation as often prefer by users for convenience reasons, since the horizontal area provided for seeker-bar navigation is very small. However, as smartphone users record and store more and more videos on their devices, and often want to share specific scenes, we need more efficient navigation means instead of pure playback features with cumbersome navigation support.

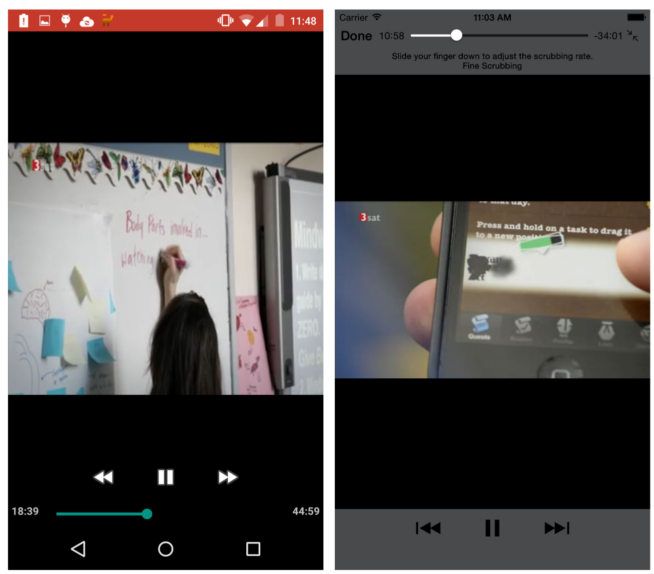

Fig. 1. Default video payer interface on smartphones. **Left**: Android 5, **Right**: iOS 8

In this paper we propose a new interaction concept for video content navigation on touchscreens and test it with large-screen smartphones. This navigation interface, called *Scrubbing Wheel*, provides superior navigation performance and higher user experience than the commonly used seeker-bar. The idea of the interface is based on the findings of [1] and inspired by the hardware design of classic Apple iPod de-

Preprint of the paper to appear in the IEEE Proceedings of the International Symposium on Multimedia 2015 (ISM 2015), Miami, Florida, USA, Dec. 2015

vices, which used a *ClickWheel* for media interaction where the user could start/pause music by clicking and browse files by performing a circular wipe gesture. We use the same interaction concept for navigation in video files through touch interaction and evaluate it with modern smartphones. The results of our user study with 24 users show that the proposed interface allows for significantly better navigation performance and is significantly less demanding in terms of user workload (e.g., mental demand, physical demand, effort, etc.), i.e. provides superior user experience.

2. RELATED WORK

In the last two decades, several proposal for improving video content navigation were made. Due to space limitations, however, here we only review works proposed for mobile devices, a topic addressed by only a few authors.

Hürst et al. proposed the *Mobile ZoomSlider* interface [2] for stylus-based navigation on handheld devices. The basic idea is a virtual seeker-bar that can be used at any screen position. The vertical click position is utilized as a parameter for navigation resolution. Moving the stylus left or right results in backward or forward navigation and the vertical position of the drag operation defines how fast the navigation action is performed. The same interaction concept has been proposed for video content navigation on PDAs and smartphones in a follow-up work [3]. The authors further suggested circular navigation gestures for stylus-based navigation in video and presented a few different interaction concepts in [1]. However, unfortunately in the evaluation no direct comparison to seeker-bar navigation was performed.

Karrer et al. [4] proposed the *PocketDRAGON* interface for touch-based video content navigation on mobile devices. Instead of an overlaid seeker-bar, they suggest direct manipulation of objects in the scene. A similar concept was already proposed by Dragicevic for video navigation on desktop PCs in an earlier work [5]. In their work, motion tracking is performed and object motion is used as basis for the dragging operation along a motion trajectory. Additionally to this object-based navigation mode, which rather improves the navigation accuracy and typically does not allow quicker navigation over longer segments in the video, they also support two-finger gestures. A horizontal wipe gesture with two fingers allows to jump to the previous or next scene in the video. However, as no evaluation has been performed by the authors, it remains unclear how well this navigation concept supports navigation tasks in videos.

Huber et al. [6] focused on improving navigation in e-learning videos rather than entertaining videos and proposed the *Wipe'n'Watch* interface. Instead of a seeker-bar, they suggest to use wipe gestures for touch-based navigation on smartphones. Their interface that operates in portrait mode is subdivided into two areas: (i) the upper area shows the actual video content and (ii) the lower area shows an overview of all available keyframes, i.e., available 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 [7]. Hence, 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. However, in their evaluation they did not directly compare to seeker-bar navigation. Therefore, it is not clear if this interaction concept is superior than a seeker-bar.

In a recent work that builds on the idea of [6] and [2], we also proposed to use wipe gestures for navigation in videos on mobile devices instead of navigation with seeker-bars [8]. The idea in this work was to use simple wipe gestures for navigation in videos, as also used on touch-based devices for inspection of photo collections, documents, or slides of a presentation. We could show through a user study with 24 participants that the majority of users (87.5%) prefer this kind of navigation over a common seeker-bar and achieve similar navigation performance.

Ganhör proposes an alternative method to interact with a video player on smartphones [9]. In landscape orientation, the screen is divided into four different interaction panes, which can be used in a few different modes: *Standard Browsing*, *Advanced Browsing* and *Progressive Browsing*. In *Standard Browsing* mode, a single tap on the left or right panes causes the interface to display the previous or the next frame. In *Advanced Browsing* mode the same functionality but support to slow down video playback is provided. Finally, *Progressive Browsing* mode is a combination of the two other modes. By moving their fingers up or down, users can increase or decrease the video playback speed.

Hudelist et al. [10] focus on highly ergonomic video browsing on touch devices and propose the *ThumbBrowser* interface that lets users browse videos with a focus on sustaining comfortable hand postures. It is assumed that the tablet is held with both hands at the sides in landscape orientation and do not show any interaction elements during playback to reduce screen clutter. Instead, interaction elements are displayed on-demand when the user touches the screen with the left or right thumb. Radial menus provide navigation controls like play, pause, fast forward and fast reverse, a vertical timeline seeker ba on the right side can be used to quickly skim through the video and jump to specific positions.

In [11], we proposed the *Keyframe Navigation Tree* interface for touch-based video content navigation on tablets. They argue that their concept of three scrollable horizontal stripes with very compact keyframes can significantly outperform the seeker-bar at known-item search tasks.

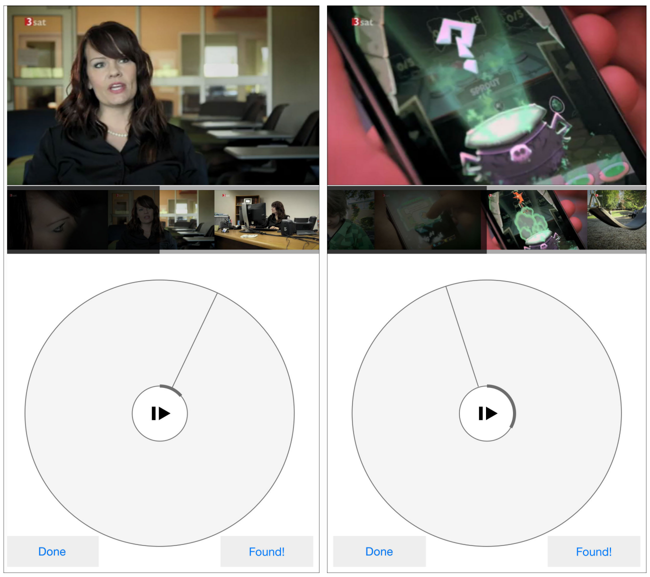

Fig. 2. *Scrubbing Wheel* interface. Video playback is controlled through touch and wipe gestures in the gray area (play, pause, fast-forward or reverse). The keyframe stripe below the video playback area scrolls together with this interaction (e.g., from left to right when wiping clockwise). A semi-transparent dark overlay of past frames, as well as a centered radial progress bar, are used for position/duration indication.

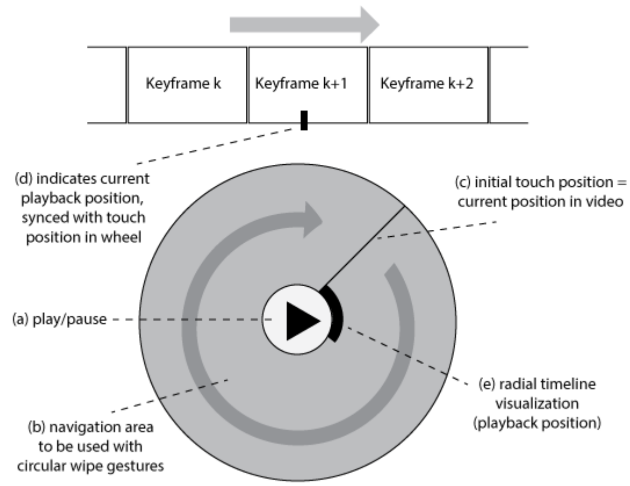

Fig. 3. Interaction with the *Scrubbing Wheel*.

3. THE SCRUBBING WHEEL INTERFACE

Our video navigation interface is based on the findings of [1] for stylus-based video navigation. Its interaction concept is inspired by the original *Click Wheel* hardware design of the classic Apple iPod device. As shown in Figure 3, it consists of a play/pause button in the center (a) and a radial area (b, indicated by the color gray) that can be used for content navigation. When the user taps this area a line appears that indicates the touch position (c). Now the user can wipe clockwise or counter-clockwise in order to forward or reverse the video. As the user does so, a synchronized stripe of keyframes of the video is scrolled in horizontal direction, where the current playback position is always in the horizontal center (d, see black indicator in the figure). If the user needs fine-grained navigation s/he performs the wipe gesture in the outer circle of the navigation area. Similarly, for coarse-grained navigation s/he performs it in the inner circle, close to the play/pause button. In the center area of the interface a radial progress bar visualizes the current playback position (e).

In our implementation we configured the Scrubbing Wheel in such a way that one complete circular wipe corresponds to about ten minutes in the video, according to empirical observations from initial user experiments. Moreover, as keyframes we use temporally subsampled frames from the video, with an equal distance of 10 seconds. Hence, a complete circular wipe gesture scrolls a horizontal list of 60 frames. Moreover, the user does only need to start the wipe gesture in the gray circular area but can actually use the entire screen for further gesture interaction. This makes use of the Scrubbing Wheel more convenient and improves user experience. Since our implementation is optimized for portrait orientation, we arrange the video playback area atop of the keyframe stripe (see Figure 2).

The flexible navigation resolution together with the synchronized scrolling of keyframes makes the Scrubbing Wheel an easy-to-use and efficient navigation tool. Though our implementation is for smartphones, it is easily portable to tablets and other touchscreen devices.

4. EVALUATION

To evaluate the Scrubbing Wheel interface, we performed a user study with 24 users (mean age 28.58, s.d. 4.77) and video content navigation tasks in a few different videos.

4.1. User Study

The 24 participants had to perform content search tasks in four different documentary videos; the tasks were based on the content of these videos:

- Video 1: Documentary about growing fruits with a duration of 25 minutes. Task: *"Find all scenes showing bananas."* (9 target positions).

- Video 2: Documentary about multi-cultural societies with a duration of 30 minutes. Task: *"Find all scenes showing a person wearing glasses."* (37 target positions).

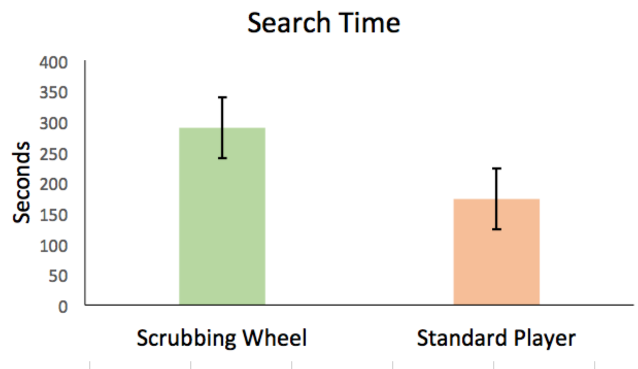

Fig. 4. Task solve time per interface (error bars: 95% confidence interval).

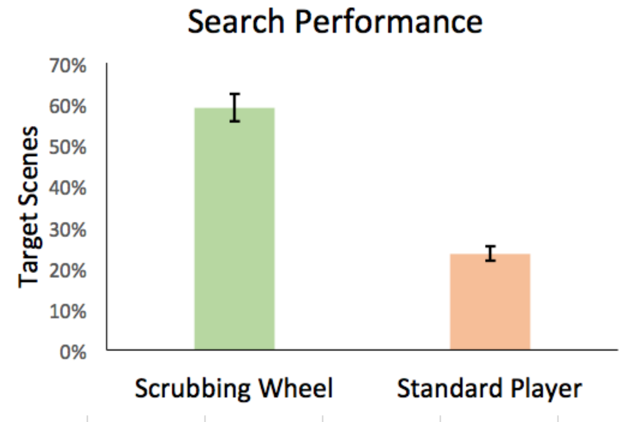

Fig. 5. Amount of found target scenes per interface (error bars: $\pm$ s.e. of the mean).

- Video 3: Documentary about planets/outer space with a duration of 35 minutes. Task: *"Find all scenes showing the Earth."* (28 target positions).
- Video 4: Documentary about 'Gamification' with a duration of 40 minutes. Task: *"Find all scenes showing smartphones."* (10 target positions).

Each participant had to perform two search tasks with the Scrubbing Wheel interface and another two tasks with the common video player interface on iOS. The test sequence of the four videos and the selection of the interface was permuted over the users, according to latin-square principle to avoid familiarization effects. Our implementation provided a dedicated button (*"Found!"*) to mark target scenes and another button (*"Done"*) to finish the current task in case no more scenes could be retrieved (see Figure 2). A test session for a user lasted for about 20-30 minutes for both interfaces. After testing one interface, the user had to complete a questionnaire with workload ratings about the interface, according to the NASA Task-Load-Index (TLX) [12]. Furthermore, after finishing all tasks the participant had to state his/her preferred interface to solve such navigational tasks.

4.2. Evaluation Results

We first evaluate the performance in terms of retrieved target scenes, for which ground truth was manually annotated. To do so we compute the average percentage of found scenes for each user and interface and perform a statistical analysis on the data. According to a *dependent t-test* the Scrubbing Wheel interface enabled users to achieve a significantly higher performance than the standard video player ($t(23) = 9.949, p < 0.0005$). The average percentage of found scenes with the Scrubbing Wheel was $59\% \pm 16\%$, whereas for the video player it was only $24\% \pm 9\%$ (see Figure 5).

When analyzing the spent search time for each interface, we can see that users also spent significantly more time with the Scrubbing Wheel interface ($t(23) = 5.149, p < 0.0005$). In average users required 290 ± 113 seconds to solve a search task with the Scrubbing Wheel, while they required only 173 ± 70 seconds with the standard video player (see Figure 4). In combination with the achieved performance, we can conclude that the reason for this is the lower suitability of the standard video player to efficiently locate target scenes in the video. Most users performed rather coarse navigation and simply overlooked a lot of relevant scenes, since they jumped back and forth in the video (see Figure 6). Hence, they typically also stopped earlier than with the Scrubbing Wheel.

The NASA Task-Load-Index [12] ratings of the test participants for both interfaces are given in Figure 7. A statistical analysis with *non-parametric Wilcoxon signed-rank tests* showed that the Scrubbing Wheel is perceived as significantly better in all seven categories (see Figure 7). More specifically, all test participants found that the Scrubbing Wheel interface better supports target search in videos (*"Support"* in the figure, $Z = -4.220,\ p < 0.0005$), that it is less frustrating than the common video player ($Z = -4.029,\ p < 0.0005$), and that it is more fun to use ($Z = -4.123,\ p < 0.0005$). Further, the vast majority found that the Scrubbing Wheel requires less effort ($Z = -3.961,\ p < 0.0005$), less mental demand ($Z = -2.532,\ p = 0.011$), less physical demand ($Z = -2.525,\ p = 0.012$), and produces less temporal pressure ($Z = -2.197,\ p = 0.028$).

At the end of the user test we asked the participants the give a vote for the preferred interface to solve the tested search tasks. It turned out that all users (100%) voted for the Scrubbing Wheel interface. When asked for the reasons of their vote, most users liked that (1) *"the Scrubbing Wheel provides much better direct navigation speed control"*, (2) that it *"allows to navigate in both fine-grained and coarse-grained manner"*, and (3) that it *"is much simpler to use because*

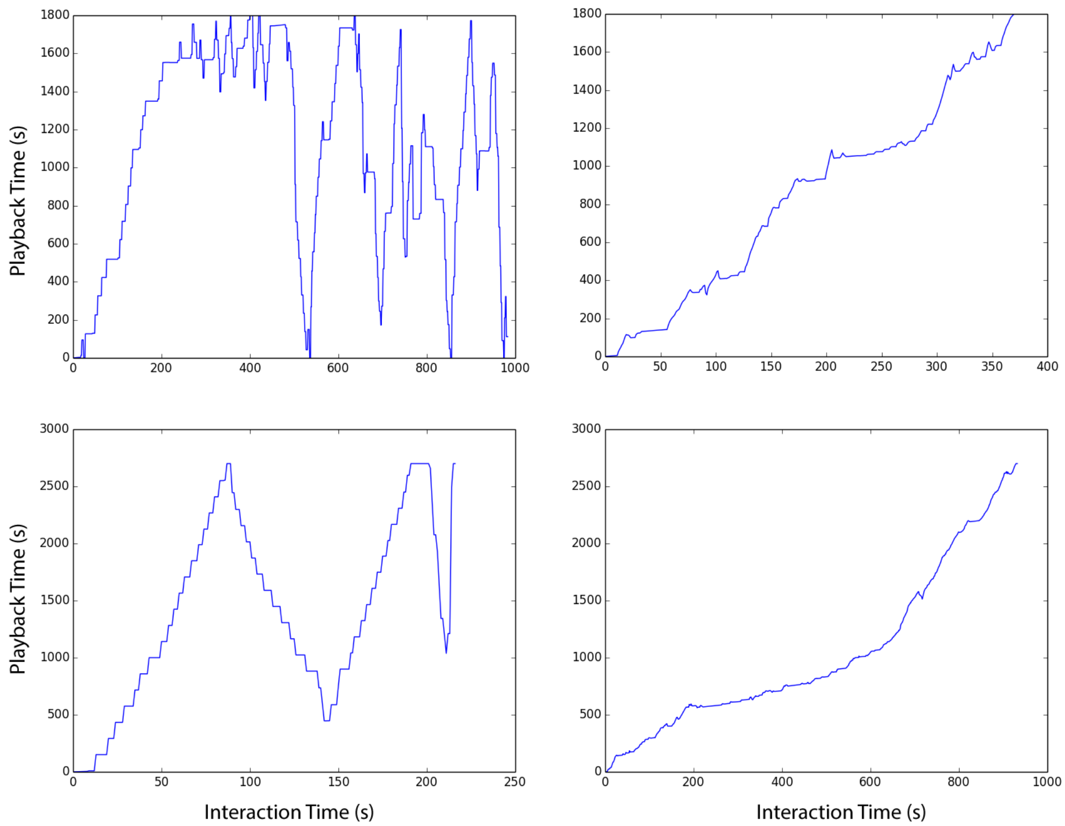

Fig. 6. Navigation behavior examples: visited positions over time with the default video player (**left**) and with the *Scrubbing-Wheel* (**right**) for two of the tested video files (one per line).

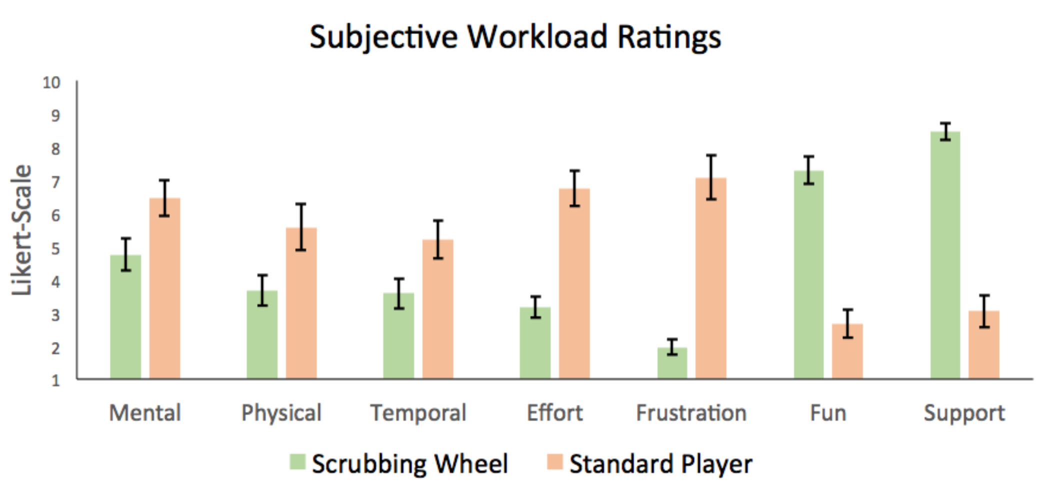

Fig. 7. Perceived workload ratings (according to NASA Task-Load-Index [12], with Likert-scale 1-10) for both interfaces (error bars: ± s.e. of the mean)

circular touch interaction is much more convenient". They also mentioned that it is *"particularly useful to have a circular navigation area together with synchronized horizontally scrolling keyframe stripe"*. In fact, this scrolling keyframe stripe alleviates the rather slow random access performance in videos on the smartphone, since the frames for the keyframe stripe are extracted simultaneously.

Therefore, we can conclude that video content navigation with the Scrubbing Wheel interface is significantly more accurate than with the default video player on iOS devices, it is significantly better in terms of subjective workload ratings, and it is the preferred interface for all 24 tested users.

5. CONCLUSIONS

We have proposed an new interaction concept for touch-based navigation in video that provides different navigation resolution and synchronized playback of keyframes. We implemented this interface on an iOS smartphone with a 5.5-inch touchscreen and tested its efficiency with 24 users in direct comparison to seeker-bar navigation with the default video player. Our results show that the Scrubbing Wheel allows for significantly higher search performance and is perceived as significantly more usable than a seeker-bar.

Acknowledgement

This work was funded by the Federal Ministry for Transport, Innovation and Technology (bmvit) and 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.

6. REFERENCES

[1] Wolfgang Hürst and Georg Götz, "Interface designs for pen-based mobile video browsing," in *Proc. of the 7th ACM Conf. on Designing Interactive Systems*, New York, NY, USA, 2008, DIS '08, pp. 395–404, ACM.

[2] Wolfgang Hürst, Georg Götz, and Martina Welte, "Interactive video browsing on mobile devices," in *Proceedings of the 15th international conference on Multimedia*, New York, NY, USA, 2007, MULTIMEDIA '07, pp. 247–256, ACM.

[3] Wolfgang Hürst and Konrad Meier, "Interfaces for timeline-based mobile video browsing," in *Proceedings of the 16th ACM Int. Conf. on Multimedia*. ACM, 2008, pp. 469–478.

[4] Thorsten Karrer, Moritz Wittenhagen, and Jan Borchers, "Pocketdragon: a direct manipulation video navigation interface for mobile devices," in *Proceedings of the 11th Int. Conf. on Human-Computer Interaction with Mobile Devices and Services*, New York, NY, USA, 2009, MobileHCI '09, pp. 47:1–47:3, ACM.

[5] Pierre Dragicevic, Gonzalo Ramos, Jacobo Bibliowitcz, Derek Nowrouzezahrai, Ravin Balakrishnan, and Karan Singh, "Video browsing by direct manipulation," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2008, CHI '08, pp. 237–246, ACM.

[6] Jochen Huber, Jürgen Steimle, Roman Lissermann, Simon Olberding, and Max Mühlhäuser, "Wipe'n'watch: spatial interaction techniques for interrelated video collections on mobile devices," in *Proceedings of the 24th BCS Interaction Specialist Group Conference*, Swinton, UK, UK, 2010, BCS '10, pp. 423–427, British Computer Society.

[7] O. de Rooij, C. G. M. Snoek, and M. Worring, "Mediamill: semantic video search using the rotorbrowser," in *Proc. of the ACM Int. Conf. on Image and video retrieval*. 2007, pp. 649–649, ACM Press.

[8] Klaus Schoeffmann, Kevin Chromik, and Laszlo Böszörmenyi, "Video navigation on tablets with multitouch gestures," in *Proceedings of The Third International Workshop on Emerging Multimedia Systems and Applications (EMSA) at the IEEE International Conference on Multimedia & Expo (ICME 2014)*. 7 2014, IEEE.

[9] Roman Ganhör, "Propane: Fast and precise video browsing on mobile phones," in *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, New York, NY, USA, 2012, MUM '12, pp. 20:1–20:8, ACM.

[10] Marco A. Hudelist, Klaus Schoeffmann, and Laszlo Boeszoermenyi, "Mobile video browsing with the thumbbrowser," in *Proceedings of the 21st ACM International Conference on Multimedia*, New York, NY, USA, 2013, MM '13, pp. 405–406, ACM.

[11] MarcoA. Hudelist, Klaus Schoeffmann, and Qing Xu, "Improving interactive known-item search in video with the keyframe navigation tree," in *MultiMedia Modeling*, Xiangjian He, Suhuai Luo, Dacheng Tao, Changsheng Xu, Jie Yang, and MuhammadAbul Hasan, Eds., vol. 8935 of *LNCS*, pp. 306–317. Springer International Publishing, 2015.

[12] S.G. Hart and L. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," in *Human mental workload*. 1988, pp. 139–183, P.A. Hancock and N. Meshkati (Eds.), Elsevier.