Fast Forward with your VCR: Visualizing Single-Video Viewing Statistics for Navigation and Sharing

Abir Al-Hajri, Matthew Fong, Gregor Miller and Sidney Fels*

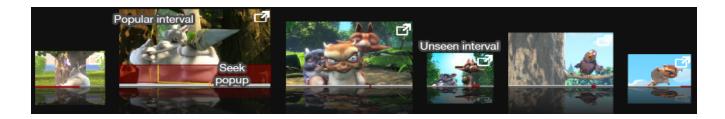Human Communication Technologies Laboratory, University of British Columbia, Vancouver, BC, Canada

Figure 1: The *View Count Record* (VCR) visualization of a single video to highlight popular or unseen content using viewing statistics derived from personal or crowd-sourced histories of video consumption. The most-viewed intervals are represented by large thumbnails, with three smaller sizes to represent less-viewed content (the smallest for unseen content). The VCR is designed to be useful for fast navigation, search via seeking (*second thumbnail from left*), instant video previews (without interrupting other content) and sharing directly from the interface.[1]

ABSTRACT

Online video viewing has seen explosive growth, yet simple tools to facilitate navigation and sharing of the large video space have not kept pace. We propose the use of single-video viewing statistics as the basis for a visualization of video called the *View Count Record* (VCR). Our novel visualization utilizes variable-sized thumbnails to represent the popularity (or affectiveness) of video intervals, and provides simple mechanisms for fast navigation, informed search, video previews, simple sharing and summarization. The viewing statistics are generated from an individual's video consumption, or crowd-sourced from many people watching the same video; both provide different scenarios for application (e.g. implicit tagging of interesting events for an individual, and quickly navigating to others' most-viewed scenes for crowd-sourced). A comparative user study evaluates the effectiveness of the *VCR* by asking participants to share previously-seen affective parts within videos. Experimental results demonstrate that the *VCR* outperforms the state-of-the-art in a search task, and has been welcomed as a recommendation tool for clips within videos (using crowd-sourced statistics). It is perceived by participants as effective, intuitive and strongly preferred to current methods.

Index Terms: H.5.2. [Information Interfaces and Presentation]: User Interfaces—; H.1.2. [Models and Principles]: User/Machine Systems—

1 INTRODUCTION

Consuming video online, on mobile devices or on home computers is now a well-accepted form of communication and entertainment, shown by the rapid growth of various providers such as YouTube[TM]. Despite the volume of video available, methods for efficient navigation and sharing of in-video content have not provided users with the ease of use or level of personalization re-

quired to accommodate their needs. Constraints such as limiting the length of videos (e.g. six seconds on Vine[TM] and 15 seconds on Instagram[TM]) can simplify the problem, however these do not address the challenges with unconstrained video.

Part of the problem is that the 3D spatio-temporal representation of video complicates relatively simple actions such as search or selection. Video search often taxes human memory by requiring memorization of a large quantity of previously-seen content. In particular, finding and selecting interesting parts has poor navigation and search support. We propose that the addition of a single-video visualization mechanism using viewing statistics will overcome some of these difficulties.

We investigate the usefulness of visualizing prior viewing by either single or multiple users to support fast navigation (to popular or unseen parts), search and directly previewing content, without interrupting normal playback. We envision users will watch videos differently when they have a visualization of their personal navigation: they can implicitly tag segments of video by re-viewing (thereby increasing the view count); it would also capture their natural behaviour, such as watching a funny section multiple times in a lengthy video. This non-linear viewing behaviour is already evident, such as in YouTube audience retention graphs[2]: videos have peaks in the graphs, implying users watch different content and likely seek to find interesting parts (unfortunately these graphs are not generally public, and require voluntary publication by video owners). Viewing graphs often show a shallow negative exponential curve (i.e cold-start problem) from crowd-sourced data, which can be very simply filtered to highlight the most popular content. Likewise, viewing statistics can be used to filter out videos where only the first few seconds are watched.

2 RELATED WORK

Researchers have proposed various navigation techniques to simplify access and improve efficiency. Simple linear video navigation

*{abira, mfong, gregor, ssfels}@ece.ubc.ca

[2]https://support.google.com/youtube/answer/1715160

Figure 2: The VCR is a new navigation tool based on viewing statistics which intuitively visualizes seen and unseen content. Each viewed interval is recorded and added to the view count for visualization. Playback is controlled either with the usual tools (play, pause, seek), or by using the VCR. For unwatched video, the view count is equal everywhere and the VCR gracefully becomes the familiar Filmstrip (Figure 3a).

can be accomplished through representative thumbnails (e.g. a film-strip metaphor), such that selecting a thumbnail directly positions the main video at the specific time corresponding to that thumb-nail [10]. *Swifter* claimed faster video navigation by displaying a grid of thumbnails instead of a single thumbnail when scrubbing a video [15]. Map-based storyboards emerged as a video navigation tool, where different intervals are directly visualized on a map [19]. Davis [8] proposed an iconic video representation language of human actions in video. Using these iconic representation allow users to visualize, browse, annotate, and retrieve video content. Others proposed playback mechanisms to help users rapidly skim through uninteresting content and watch interesting parts at a normal pace [4]. The timeline can also be controlled by directly manipulating video content (i.e. using object motion) [9].
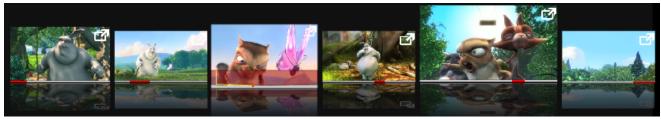
Video summarization aims to shorten videos to emphasize the important content and reduce time needed for viewing. For instance, Correa et al. proposed a system for generating dynamic narratives from videos where a compact, coherent and interactive poster is created for each video [6]. They use a series of spatial-temporal masks to improve the output quality of stitching fore-ground and background regions of video frames. Daniel et al. [7] and Nguyen et al. [18] applied computer vision algorithms to develop a 3D volume-based interface for video summarization and navigation. It enabled users to understand the content and easily navigate to the intended parts of the video. Video summarization is represented by a large body of work [1, 17, 22], however, in general these methods do not take advantage of any implicit information

gathered as users consume video, which may be used to personalize the user experience.

Video navigation histories and social navigation provide potential interaction techniques for fast navigation, event search and video summarization. People in real scenarios rewatch parts of videos that are important, interesting, affective, or hard to understand [20]. This can be seen most simply in YouTube's or Vimeo's feature for sharing a video from a specific start time, providing an exact use case of what we propose. This watching behaviour leaves digital footprints on the video frames creating a non-flat video histogram emphasizing the interest of each part of the video. Researchers have effectively employed this data for different purposes. For example, Yu et al. have used video browsing histories to rank different scenes within a video, then use the rankings to extract clips to generate video summaries [24]. Mertens et al. have used the video timeline itself to visualize users' viewing 'footprints' using different brightness levels, which lets users quickly navigate to the most viewed scenes [16]. However, this does not supply a visualization of the video, which inhibits search (when searching for a previously seen event, users need to remember its approximate location), and intervals are not defined as a whole which prevents sharing directly from the visualization. Al-Hajri et al. offered users access to their video browsing histories by visualizing the intervals a user watched as a separate thumbnail [2, 3]. Users were able to find previously-seen content more quickly than when their history was not available. However, rewatched intervals were visualized multiple times (which increased the search space) which may ham-

(a) Filmstrip (equivalent to VCR without visualization of viewing statistics)

(b) View Count Record (VCR)

Figure 3: The View Count Record (VCR) component visualizes the video viewing statistics. When no viewing history is available, the VCR presents a familiar filmstrip (a). When a history is available, our view count manipulation algorithms can be applied to visualize popular intervals within the video, leading to fast personal navigation and social navigation applications. Each thumbnail can be searched via seeking in the popup red time bar when hovering the cursor over the preview image.

per the search task.

Video navigation history can play an important role in user-based information retrieval from videos on the web. Shamma et al. [21] and Yew et al. [23] have proposed a shift from content-based techniques to user-based analysis because it provides a more promising basis for indexing media content in ways that satisfy user needs. Leftheriotis et al. used viewing statistics as a tool for extracting a representative image or thumbnail based on the most-viewed frame within a video [14]. Hwang et al. applied the viewing statistics to place the most-viewed video content in provider networks rather than using the complete video to reduce their storage and network utilization [13]. User-based viewing statistics were demonstrated to be at least as effective at detecting video events as when content-based anaylsis techniques (i.e. computer vision) were used [12]. Fong et al. [11] employed users' navigation behaviour to propose a new casual video editing interaction technique. They showed that using their technique participants took at most two-thirds the time taken by the conventional method to select video segments and create playlists.

3 SINGLE VIDEO NAVIGATION

The objective of our research is to design a visualization that supports fast in-video navigation (play most popular or unseen parts), search (seek within intervals with prior knowledge e.g. 'seen the event before' or 'never seen the event'), preview, and instant sharing (share a single interval directly). To accomplish this, we use viewing statistics (personal or crowd-sourced) as the basis for a modification to the well-known filmstrip visualization [5], to create the *View Count Record* (VCR). The VCR uses a variable thumbnail size (and variable interval length) to reflect the relative popularity of intervals. This is similar in concept to timeline footprints [16], however timeline footprints do not allow direct navigation (we do not require seeking), in-place preview or direct interval sharing. We used size instead of colour since we are representing intervals using thumbnails where colour discrimination may be confused with the thumbnail content and would be difficult to differentiate for some videos. We also piloted the visualization using a coloured frequency bar attached to each thumbnail to indicate the popularity; while the information was welcomed, the visualization was re-

ported as cluttered. The *VCR* applies a histogram visualization using thumbnails where the height of these thumbnails indicates the height of the histogram bars.

Whenever a segment of video is played, the video ID and timestamps for the interval's start and end are recorded. An accumulated view count is maintained for the video at a given resolution (e.g. 15 samples per second of video). The VCR, shown in Figure 3, consists of a fixed number of *video segments*[3] (described below). The duration and size of each segment is based on how often its corresponding interval has been viewed. If no viewing statistics are available, the VCR appears as a normal filmstrip as shown in Figure 3(a).

3.1 VCR Construction

The construction of the *VCR* starts by gathering intervals of time in which consecutive frames have equal view counts. While there are intervals less than a set threshold, the algorithm attempts to merge these intervals with one of their neighbouring intervals. The neighbour to merge with is determined by two criteria: first by the difference in view counts, and if the difference of view counts are equal, then by the duration of the neighbouring intervals. The merging process chooses the smallest difference in view counts, or the smallest interval duration. This process repeats until all intervals' duration are greater than the preset threshold.

Upon the completion of the merging process, the *VCR* contains a set of intervals with duration that are greater than the preset threshold. However, since the number of items in the visualization component is limited, we must reduce the set of intervals to match. Thus we look at the peaks of the view count graph, keep the highest peaks, and merge the other intervals until we get to the desired resolution. Conversely, if we do not have enough intervals, we linearly sample and split intervals until we have enough.

We then create our visualization by using a Video Segment component (described below) for each interval. The size of each segment is based on a ratio of the current segment's view count to the maximum view count for the video. The *VCR* updates automatically when the video is paused, based on the latest viewing statis-

---

[3]In our interface, we used 6 segments based on the width of the interface and the maximum width of a single *VCR* segment.
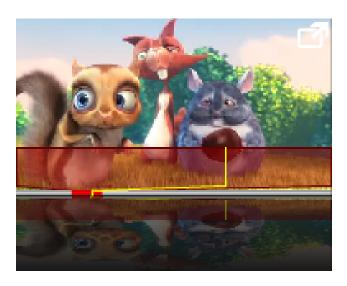
Figure 4: Each video thumbnail in the VCR is visualized as small video segments. Each segment is seekable and playable on mouse events. The red/gray portion at the bottom of the widget indicates the temporal location of its interval within the complete video. The yellow line illustrates the current seeking point within the thumbnail, within the zoomed interval for higher-resolution seeking.

tics (it does not update while viewing so as not to distract from the main video). It illustrates to the user how they consumed the video and which parts were viewed the most/least. This provides a simple mechanism to find or navigate back to these segments when needed.

3.2 VCR Visualization

Each interval (i.e. *video segment*) in the *VCR* is represented as a thumbnail displaying the first frame. The temporal location of the interval with respect to the video is visualized as a red line on a grey background beneath the thumbnail (the location bar) to help users spatially contextualize the temporal location of intervals within the complete video. Video segments support: seeking using a popup overlay which expands from the red line in the location bar, and the thumbnail's image updates to reflect the seeked content (see Figure 4); playing the video directly, without interrupting the main player; playing the video in the main player, either by clicking the top right arrow or dragging the interval to the main player (dragging with the seek bar active plays the interval from that specific time); and finally, sharing of the entire interval by drag-and-drop.

3.3 VCR Scalability

The *VCR* is not affected by the length or duration of the video being visualized since the algorithm as described in Section 3.1, merges (or linearly samples and splits intervals) until the *VCR* gets to the desired resolution (i.e the required number of segments). However, due to the limited space and the fixed number of video segments, some medium-height peaks may be diminished and not easily viewed in the *VCR*. To alleviate this problem, the interface supports a zoom feature (via mouse wheel) where the selected video segment expands and is represented by its own *VCR* with the same number of segments. When a segment is zoomed-in, the *VCR* updates to visualize segments within that zoomed segment only and hides any other segments. Thus the *VCR* always uses the same number of video segments.

4 EVALUATION

We designed a comparative study to investigate: 1) if our visualization of video navigation provides faster search for user-specific af-

fective intervals, and if users prefer our visualization for this task; 2) if crowd-sourced histories provide good summaries of video. Subjects were asked to find and share their favourite intervals using either the *VCR* or *Filmstrip* visualizations. We compared against the filmstrip design instead of timeline footprints [16] for several reasons: footprints does not easily let a user directly select or share a full interval; video cannot be previewed inside the footprints visualization (*VCR* and *filmstrip* can both directly preview without seeking); *VCR* and footprints could be used together, so we believe a comparison against *Filmstrip* is more informative.

4.1 Participants

Ten paid volunteers, 6 female and 4 male, participated in the experiment. Participants ranged in ages from 19 to 35. Three of the participants were undergraduate students while the rest (i.e. 7 participants) were from the general public (non-academic). Each participant worked on the task individually. All participants were experienced computer users and have normal or corrected to normal vision. Seven participants watch online videos on a daily basis and the other three watch videos 3-5 times a week. Five of the participants watch 1-3 videos on average per day, while three watch 3-5 videos per day and two watch more than 10 videos per day on average.

4.2 Design and Procedure

Two different navigation modes were tested: *Filmstrip* and *VCR*. The case where no history was available was represented by the state-of-the-art *Filmstrip*, shown in Figure 3(a). Each participant tried both modes to navigate and share their preferred parts of the video. Participants were divided equally into two groups where Group 1 used *Filmstrip* first and *VCR* second, and Group 2 had the order reversed. Participants freely watched a set of 5 different videos (Disney short animations) between 3 and 5 minutes long. Video length does not affect the *VCR* as mentioned in Section 3.3, however, due to the time constraints of the experiment short videos were tested.

To ensure all participants had seen equivalent video, the tasks began after all videos were viewed. For each video, participants were asked to list five intervals they would like to share; these were recorded by the researcher. The researcher chose an event from those provided by the participant which they must find: the search task began by clicking 'Find', choosing a video from a grid of thumbnails, and then the navigation layout for the current mode was displayed - the participant used this to find an interval representing the event. The interval is submitted for consideration by playing it: if approved by the researcher as correct, the task is complete.

Each participant performed a total of 14 search tasks (2 modes × 7 intervals); they were asked to perform as quickly as possible. For each task, the completion time, the number of previews and the number of zoom events were recorded. The completion time was measured from when the participant clicked on a 'Find' button until the moment they found the correct interval (confirmed by the researcher). The navigation behaviour and statistics were recorded during the viewing phase. The participants were also asked to rank each mode based on speed, ease and preference.

Upon the completion of the sharing tasks, participants started the second task where they were shown a short version of each video, automatically created from crowd-sourced histories (described below). Participants were asked if they thought the shortened version was a good summarization and whether each segment in the crowd-sourced version matched their own affective segments; the experiment ended when participants had ranked all 5 shortened videos and their corresponding segments. The final task was to fill out a questionnaire to rank the modes, and provide feedback on the interface, its features and their experience. The experiment lasted

Table 1: Results of the comparative study for the interval retrieval task, showing a significant advantage using our method (VCR) in terms of completion time. Note: SD = standard deviation; completion time measured in seconds. * $p < 0.03$

	Filmstrip		VCR		
	Mean	SD	Mean	SD	t-test
Completion Time	24.31	10.42	21.11	5.38	-2.28*
No. of Previews	40.39	35.41	35.53	27.56	0.88

approximately one hour per participant.

4.3 Crowd-Sourced Data Collection

Six graduate students (2 female, 4 male, aged 24 to 37) completely separate from participants in this study, voluntarily participated in the crowd-sourced data collection. Participants were invited prior to the experiment to freely watch and navigate the same set of videos while their viewing statistics were recorded. Their data was then aggregated and visualized using the *VCR*. At least 9 peaks existed for each video. However, due to the experiment time constraints (one hour), we decided to use only the highest 5 peaks of each video in the shortened videos that were tested.

4.4 Results and Discussions

Most participants commented that they enjoyed their time using the interface and they can imagine seeing its features applied, especially in social networking websites. They foresaw its applicability as a navigation aid for un-watched videos where social navigation can be leveraged for the benefit of future viewers, as well as a summarization tool for their own videos. Participants were impressed by how closely the crowd-sourced popular intervals matched their own preferences for best intervals, confirming that in most cases this would provide an effective tool for navigating new video.

4.4.1 Search Task

The main task in the experiment was to search for previously-seen preferred intervals: each participant was able to complete each search task in less than one minute (for all 14 trials). A paired-samples t-test analysis determined the significance of the results in terms of the average completion time per search and the average number of previews per search. The analysis, shown in Table 1, demonstrated that the search task using *Filmstrip* took significantly more time than with the *VCR*. Participants were asked to rank the different modes for preference, ease and speed: they ranked the *VCR* as the most liked, easiest and fastest mode, which coincides with the quantitative results. This indicates that having access to the user's personal navigation record is useful for finding previously-seen content within video, and that our visualization cues (e.g. size) of the mostly watched segments helped users to quickly and easily navigate to the correct intervals.

In terms of the average number of previews, the results revealed no significant difference between the two modes, which we did not anticipate. This can be due to the fact that many view count peaks can exist within a single video segment of the *VCR*, and that some segments ended up much smaller in size which made it harder to navigate. When analyzing the participants' navigation history, we found that participants created 11 history segments on average per video. This means that when using heuristics some *VCR* segments had more than one peak since there are only 6 segments in the *VCR*. However, as we mentioned in Section 3.3, we added the zoom functionality to mitigate this. Participants rarely used the zoom feature and preferred to navigate through these segments instead which explains the large number of previews.

4.4.2 Agreement With Crowd-Source

All participants agreed that the shortened video (created automatically using the crowd-sourced information) was an effective summary of the video content. Before using the interface, participants were asked whether they would use others' recommendations as a tool for navigating unseen videos; we were most interested in discovering if participants views would change after using our interface. Most said they would not use recommendations, however after using the interface and viewing the shortened video they expressed surprise at the quality of the summary. Participants mentioned that having the crowd-sourced *VCR* would save time, especially for long videos, since they can decide whether to watch the entire video or just the summary, or even just parts of the summary.

For each video, participants were asked to rank each segment derived from the crowd-sourced data. At least 8 participants out of 10 agreed that each segment represented something they liked or illustrated an affective clip. Out of a total of 25 segments, 7 segments were liked by 8 participants, 8 segments were liked by 9 subjects, while the remaining 8 were liked by all participants. The negative ranking of segments by some participants was due either to religious beliefs or perceived violent content, while other participants considered these segments to be funny. We expected the variation between participants, however, we did not predict the generally high level of agreement. This suggests implicit tagging of video from many users may serve as a valuable navigation tool for online video.

4.4.3 Ranking of Visualization's Features

From the aggregated results of the questionnaire (measuring ease-of-use and usefulness), the average ranking across all components and features was 5.82 out of 7. All features were ranked above 5 except for three items which were: getting started ($(M = 4.5)$, remembering how to use the interface ($M = 4.6$), and using the zoom ($M = 4.3$). The zoom scored slightly lower due to the mouse wheel sensitivity being reported as too high, which led to some participants becoming confused or frustrated. This could also explain the low usage of this feature while performing the tasks where only 2 participants used it for 4 tasks out of 140 tasks (10 participants $\times$ 2 modes $\times$ 7 tasks) when searching for events. This has been taken into account for future versions of the interface. Overall participants appreciated the zoom since it enabled them to get a more detailed view of the video's content.

4.4.4 Participants' Feedback

There were some overwhelming impressions and comments made by the participants about the interface. One participant commented that "I definitely see how this would be really helpful for long videos because I will not have to waste my time watching the whole video again to get to the important stuff. I could directly use my previous history to navigate to these intervals." Others said "I would love to see this implemented within social websites. I could see how it would save my time when viewing new videos"; "It is really cool and easy to use. When are you going to apply this to online video websites?"; and finally "I didn't expect others' history would be useful, but, you showed me it is."

5 CONCLUSION

We have presented a new way to visualize and navigate a video space using the View Count Record (*VCR*), that provides simple navigation, search, preview and sharing of video intervals. Our comparative study based on a use case of searching and sharing found significant quantitative results in favour of our method, as well as being positively perceived by the participants. Using crowd-sourced data as a tool for recommending segments within videos (i.e. social navigation) was found to be appreciated, and we confirmed that the summaries generated from crowd-popular segments were effective at communicating the content of video. The *VCR* was rated highly by users who recommend integrating this mechanism into online video websites.

We will investigate how this navigation mechanism may be extended to multiple videos, to provide users with an intuitive and fast navigation mechanism for their video collection, as well as for serendipitous discovery of new video based on crowd-sourced information. We intend to explore how users respond to these mechanisms, in conjunction with the presented *VCR*, via a field study utilizing online video; extensive data will help determine general users' current viewing behaviour for all types of video, and how it changes when given a *VCR* and other methods based on viewing statistics.

REFERENCES

[1] M. Ajmal, M. Ashraf, M. Shakir, Y. Abbas, and F. Shah. Video summarization: Techniques and classification. In L. Bolc, R. Tadeusiewicz, L. Chmielewski, and K. Wojciechowski, editors, *Computer Vision and Graphics*, volume 7594 of *Lecture Notes in Computer Science*, pages 1–13. Springer Berlin Heidelberg, 2012.

[2] A. Al-Hajri, G. Miller, S. Fels, and M. Fong. Video navigation with a personal viewing history. In *Human-Computer Interaction – INTERACT 2013*, volume 8119 of *Lecture Notes in Computer Science*, pages 352–369. Springer Berlin Heidelberg, 2013.

[3] A. Al-Hajri, G. Miller, M. Fong, and S. Fels. Visualization of personal history for video navigation. In *Proceedings of the ACM CHI Conference on Human Factors on Computing Systems*, CHI'14, New York City, New York, U.S.A., 2014. ACM.

[4] K.-Y. Cheng, S.-J. Luo, B.-Y. Chen, and H.-H. Chu. Smartplayer: user-centric video fast-forwarding. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, pages 789–798. ACM, 2009.

[5] M. Christel and N. Moraveji. Finding the right shots: assessing usability and performance of a digital video library interface. In *Proceedings of the 12th annual ACM international conference on Multimedia*, MULTIMEDIA '04, pages 732–739. ACM, 2004.

[6] C. D. Correa and K.-L. Ma. Dynamic video narratives. In *ACM SIGGRAPH 2010 Papers*, SIGGRAPH '10, pages 88:1–88:9, New York, NY, USA, 2010. ACM.

[7] G. Daniel and M. Chen. Video visualization. In *Proceedings of the 14th IEEE Visualization 2003 (VIS'03)*, VIS '03, pages 54–, Washington, DC, USA, 2003. IEEE Computer Society.

[8] M. Davis. Human-computer interaction. chapter Media Streams: An Iconic Visual Language for Video Representation, pages 854–866. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1995.

[9] P. Dragicevic, G. Ramos, J. Bibliowitcz, D. Nowrouzezahrai, R. Balakrishnan, and K. Singh. Video browsing by direct manipulation. In *Proceedings of the Conference on Human Factors in Computing Systems*, pages 237–246. ACM, 2008.

[10] S. M. Drucker, A. Glatzer, S. De Mar, and C. Wong. Smartskip: consumer level browsing and skipping of digital video content. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves*, CHI '02, pages 219–226. ACM, 2002.

[11] M. Fong, A. Al-Hajri, G. Miller, and S. Fels. Casual authoring using a video navigation history. In *Graphics Interface*, GI '14. Canadian Information Processing Society / ACM, May 2014.

[12] C. Gkonela and K. Chorianopoulos. Videoskip: event detection in social web videos with an implicit user heuristic. *Multimedia Tools and Applications*, pages 1–14, 2012. 10.1007/s11042-012-1016-1.

[13] K.-W. Hwang, D. Applegate, A. Archer, V. Gopalakrishnan, S. Lee, V. Misra, K. Ramakrishnan, and D. Swayne. Leveraging video viewing patterns for optimal content placement. In R. Bestak, L. Kencl, L. Li, J. Widmer, and H. Yin, editors, *NETWORKING 2012*, volume 7290 of *Lecture Notes in Computer Science*, pages 44–58. Springer Berlin Heidelberg, 2012.

[14] I. Leftheriotis, C. Gkonela, and K. Chorianopoulos. Efficient video indexing on the web: A system that leverages user interactions with a video player. In *Proceedings of the 2nd International Conference on User-Centric Media (UCMEDIA)*, 2012.

[15] J. Matejka, T. Grossman, and G. Fitzmaurice. Swifter: Improved online video scrubbing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 1159–1168, New York, NY, USA, 2013. ACM.

[16] R. Mertens, R. Farzan, and P. Brusilovsky. Social navigation in web lectures. In *Proceedings of the seventeenth conference on Hypertext and hypermedia*, HYPERTEXT '06, pages 41–44. ACM, 2006.

[17] A. G. Money and H. Agius. Video summarisation: A conceptual framework and survey of the state of the art. *J. Vis. Comun. Image Represent.*, 19(2):121–143, Feb. 2008.

[18] C. Nguyen, Y. Niu, and F. Liu. Video summagator: An interface for video summarization and navigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 647–650, New York, NY, USA, 2012. ACM.

[19] S. Pongnumkul, J. Wang, and M. Cohen. Creating map-based storyboards for browsing tour videos. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*, UIST '08, pages 13–22. ACM, 2008.

[20] C. A. Russell and S. J. Levy. The temporal and focal dynamics of volitional reconsumption: A phenomenological investigation of repeated hedonic experiences. *Journal of Consumer Research*, 39(2):341–359, August 2012.

[21] D. A. Shamma, R. Shaw, P. L. Shafton, and Y. Liu. Watch what i watch: using community activity to understand content. In *Proceedings of the International Workshop on multimedia information retrieval*, MIR '07, pages 275–284. ACM, 2007.

[22] B. T. Truong and S. Venkatesh. Video abstraction: A systematic review and classification. *ACM Trans. Multimedia Comput. Commun. Appl.*, 3(1), Feb. 2007.

[23] J. Yew, D. A. Shamma, and E. F. Churchill. Knowing funny: genre perception and categorization in social video sharing. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, pages 297–306. ACM, 2011.

[24] B. Yu, W.-Y. Ma, K. Nahrstedt, and H.-J. Zhang. Video summarization based on user log enhanced link analysis. In *Proceedings of the eleventh ACM international conference on Multimedia*, MULTIMEDIA '03, pages 382–391. ACM, 2003.