Panopticon: A Parallel Video Overview System

Dan Jackson, James Nicholson, Gerrit Stoeckigt, Rebecca Wrobel, Anja Thieme, Patrick Olivier
Culture Lab, Newcastle University
Newcastle upon Tyne, UK.
{dan.jackson, james.nicholson, g.stoeckigt, r.wrobel, anja.thieme, p.l.olivier}@newcastle.ac.uk

ABSTRACT

Panopticon is a video surrogate system that displays multiple sub-sequences in parallel to present a rapid overview of the entire sequence to the user. A novel, precisely animated arrangement slides thumbnails to provide a consistent spatiotemporal layout while allowing any sub-sequence of the original video to be watched without interruption. Furthermore, this output can be generated offline as a highly efficient repeated animation loop, making it suitable for resource-constrained environments, such as web-based interaction. Two versions of Panopticon were evaluated using three different types of video footage with the aim of determining the usability of the proposed system. Results demonstrated an advantage over another surrogate with surveillance footage in terms of search times and this advantage was further improved with Panopticon 2. Eye tracking data suggests that Panopticon's advantage stems from the animated timeline that users heavily rely on.

Author Keywords

Multimedia tools; video surrogates; video browsing; surveillance video.

ACM Classification Keywords

H.5.1 Multimedia Information Systems

INTRODUCTION

The volume of digital video data has grown dramatically in recent years. By the end of 2012 video accounted for half of consumer Internet traffic and, by 2016, 1.2 million minutes of video is forecast to transfer over the Internet every second [16]. Supporting applications such as video browsing, selection and editing within this information-dense resource requires interface technologies that provide efficient representations of video content. However, gaining a rapid visual overview of video data, and searching long videos for specific information, are both challenging and time consuming activities [2].

Video surrogates allow users to view a more efficient representation in lieu of the original video. Storyboard surrogates are the most popular form of video overview [1] and are typically presented in a grid layout with time increasing (not necessarily uniformly) left-to-right, top-to-bottom. This organization allows the user to see the representative frames simultaneously while maintaining their temporal relationship. The result is that storyboards are able to present users with an overview of the context, but at the expense of losing the motion information from the video.

The purpose of this paper is to describe a new video surrogate system, *Panopticon*, and compare it with other commonly implemented systems in terms of users' seeking experience and gist comprehension. Panopticon is distinctive in that it does not remove any of the frames and therefore does not compromise on context. This is achieved through the animation of the timeline which results in an overview of the video that displays all frames within an approximate 10 second time window. Based on this maximization of contextual information to aid visual search, we hypothesized that Panopticon would help users find information in videos faster than other surrogates. To test our assumptions we ran a series of evaluations in the form of two user studies to determine how users performed when asked to find information in previously unseen videos. The results from these studies show evidence that Panopticon's animated timeline is an improvement upon existing static systems for some categories of video – i.e. unedited and especially surveillance videos.

We first present relevant literature before introducing the two user studies, where Panopticon is compared to *VideoBoard* – a video surrogate system similar to those found in video editing applications – and a popular online video player, *YouTube*. The first study focused on highlighting the differences in search times between video types and participants' ability to understand the gist of the video. In the second study we evaluated an improved version of Panopticon based on feedback from participants. We conclude with a discussion on the differences between systems and videos and present the wider implications of this work.

BACKGROUND

The increased affordability of personal video recorders and smartphones have made the capture of personal video simple and convenient, resulting in rapidly expanding content libraries. With the rise of faster broadband internet connections to homes, YouTube has become one of the most popular internet video content libraries in the world

UIST'13, October 08 - 11 2013, St Andrews, United Kingdom
Copyright 2013 ACM 978-1-4503-2268-3/13/10…$15.00.
http://dx.doi.org/10.1145/2501988.2502038

[2]. The availability of large amounts of video has created a new problem for users [4] – the need to find specific scenes or information in videos, or the need to take in the information quickly from potentially long videos. To address such needs YouTube, for instance, introduced features that facilitate contextual information for users when seeking and searching through videos: by holding and moving the play-head at any point, the interface presents the user with time-dependent context (see Figure 1, approximately 30 seconds at any one time for a 30 minute video). However, the implementation not optimal for long videos [2].

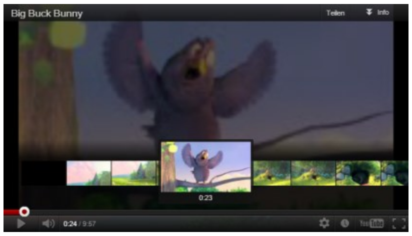

Figure 1. YouTube interface showing local context.

Video surrogates refer to an efficient representation of a video that can be used as a point of reference, for example, in the decision to view the full video or not. A variety of video surrogates are currently available – e.g. storyboards, slideshows and fast-forwards (for an overview see Shoeffmann et al. [11]). More recently, multimodal surrogates incorporating audio and visual features have been developed and evaluated (e.g. [8]). While results have been encouraging the use of audio is not always practical (e.g. in a loud office environment, or from the lack of necessary equipment). Gil et al. [5] proposed a novel text-based method for obtaining an overview of a video using tags with high participant satisfaction, but Wildemuth et al. [15] suggest using matching surrogates for media – for example text surrogates for text and video surrogates for video – in order to minimize complications in the decision-making process.

The video surrogate literature has focused on gist comprehension [3, 12] but little research has been done on the suitability of surrogates for searching and seeking information within videos. Tse et al. [13] investigated both gist comprehension and object recognition in surrogates in the context of playback speed and simultaneous displays and concluded that static simultaneous displays (such as 12 static frames representing points in the video) were better for identifying objects than a dynamic single display (such as a series of frames being played one after the other in a single position). More recently, Matejka et al. [9] proposed Swifter, an enhanced system for video scrubbing that presents scenes in a grid layout to facilitate scene identification. This technique was found to be an improvement over existing video scrubbing methods (e.g. utilized by YouTube, Netflix, etc.) and suggests that the storyboard grid layout could be beneficial for video search.

Popular video editing applications such as *iMovie* and *Final Cut Pro* utilize a 'timeline' feature where the entirety of the video is summarized in a storyboard-like approach. For the purposes of video seeking and searching, work in the area of visual search by Hollingworth [7] has shown that a preview of a scene – with either the target present or not present – improves the search times for that target over the no-preview condition. With this in mind, the use of a storyboard layout as a video surrogate is preferred, given the additional context presented by the illustration of the whole video. Work by Christel [1] confirms this preference for storyboards for video visualization.

An interesting area where surrogates may be of some use is video surveillance. In some systems, surveillance video search is done automatically, but user input on decisions is still necessary at times. For example, it is common for alerts to be automatically triggered when unusual events are spotted, but the resulting alert is received without the video footage in order to save bandwidth. It is possible to request the footage at the expense of the bandwidth but then the decision arises between downloading more than just the event for context or save resources and potentially miss important information [6]. The use of a suitable video surrogate in this situation would alleviate the bandwidth concerns while not compromising on context.

In general, existing video surrogates either present a static view of the context, losing valuable motion information, or a dynamic representation that is removed from the context of the video as a whole. In the following section, a new video surrogate system is described, Panopticon, which is designed to take advantage of context in visual search.

THE PANOPTICON SYSTEM

In the Panopticon system, a grid of thumbnails is displayed adjacent to a traditional video player (see Figure 2). The grid serves as a seek area and displays a 'timeline' where the whole video is presented in the form of thumbnails representing approximately every tenth second (depending on the video duration). Displaying frames from a video in parallel is a powerful method of providing the user with a rapid overview of the video, and by animating the thumbnails appropriately, the user can easily follow any part of the video without the need to seek.

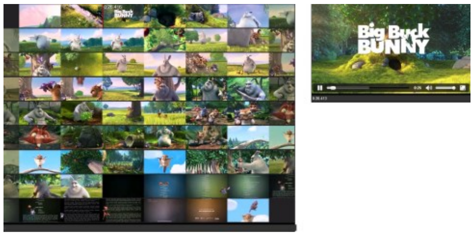

Figure 2. Panopticon interface. Left: dynamic timeline, right: video player.

Unlike other surrogates, Panopticon retains all of the frames from the original source – thereby preserving both the motion information and the overall context. Each of the thumbnails is animated in a loop by advancing the frame indices presented to the user over time. For example, the first thumbnail could play the first 10 seconds from the start of the video (0:00 until 0:10) and then restart. In parallel, the second thumbnail plays the next 10 seconds (time 0:10 to 0:20), and so on. This provides a consistent spatiotemporal layout – the opening scene of a video will always stay in the top left corner, and the end of a video will always be in the bottom right.

As each thumbnail reaches the time the next thumbnail started from (after 10 seconds) the thumbnail videos loop – for example, the first thumbnail ends at 0:10 and directly jumps back to 0:00 while the second thumbnail starts again at 0:10. To prevent this from causing a spatial discontinuity – so that a user does not have to adjust their gaze every 10 seconds when following the video – the grid is animated so that it slides smoothly to the right over time. By the time the first thumbnail has played its content, its position is where the second thumbnail originally started – when the loop occurs, the thumbnails reset to their original location, yet no visual jump is perceived. This makes it possible to smoothly follow the content at any point, without confusion.

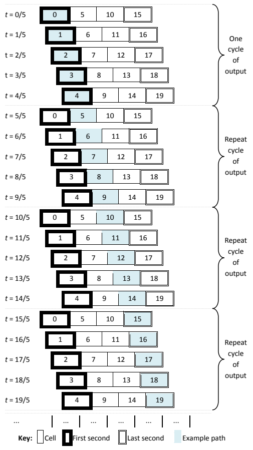

Figure 3. An example of a single-row Panopticon grid over time. Video frame indices shown at each time step, *t* (5 fps, 1 sec. cell interval, 4 sec. duration). One uninterrupted path through the video is highlighted.

Figure 3 shows an illustration of the Panopticon technique using a single-row grid for brevity. The start of the video remains on the left, the end remains on the right. An overview of all 20 frames has been given after just the first five steps, and yet any sub-sequence (even the whole video) can be followed without interruption. Moreover, it can be seen that the displayed sequence repeats in the same period as the interval between each grid cell (in this example, every five frames). These five frames, when repeatedly looped, represent exactly all that is required to be shown to the user. This fact enables the whole Panopticon overview to be computed once as a very short video loop and used in resource-constrained environments (such as web browsers). An interactive example of the Panopticon system is available online[1].

The discontinuity at the end of each line is minimized as the grid cells slowly fade out at the end of a line once they have faded in at the start of the next line, allowing the user time to adjust their gaze to the duplicate cell below.

In selecting the parameters of a Panopticon display, thumbnail sizes are maximized (while preserving aspect ratio) to produce a grid layout that provides a between-thumbnail time close to a chosen interval (e.g. 10 seconds). A minimum limit on thumbnail size (for legibility) means that, for longer videos, the interval must grow, to the detriment of easily identifying momentary actions.

VideoBoard comparison system

Based on previous work in the area of video surrogates, we predict that a static video surrogate system will facilitate information search within videos when compared with a video player such as YouTube. To this aim, we have created *VideoBoard,* a storyboard presentation with a video player (see Figure 4), which mimics commercial video editing software by providing fine-grained seek previewing of frames between thumbnail representations when moving the mouse pointer over the thumbnails.

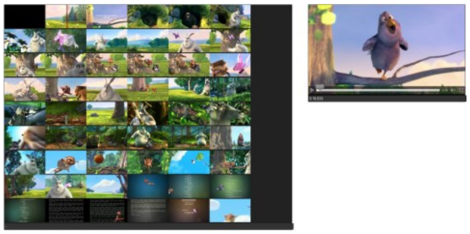

Figure 4. VideoBoard interface. Left: static timeline, right: video player.

Based on the work by Hollingworth [7], we predict that Panopticon, a system that preserves all the context of the source, will facilitate information search when compared with VideoBoard.

STUDY 1

This initial study was designed with the aim of comparing user performance in searching previously unseen videos with Panopticon and two other systems – VideoBoard and YouTube – using three different video types. Participants were evaluated in both object recognition and gist comprehension.

Design

A 3×3 factorial mixed design was used to evaluate the time taken (seconds) for participants to find information with one of three video systems: Panopticon, VideoBoard, and

[1] http://di.ncl.ac.uk/panopticon

YouTube. Participants were required to find information in three separate videos: edited footage, unedited footage and surveillance footage.

Participants

A total of 36 participants (13 female, 23 male) aged 21 – 48 years (M = 27) were recruited from the university. They were randomly assigned to one of the three video conditions (Panopticon n=12, YouTube n=12, VideoBoard n=12). Participants were provided with refreshments, but were not financially compensated.

Materials

Three video systems were used for evaluation in this series of studies: Panopticon (Figure 2), VideoBoard (Figure 4), and YouTube (Figure 1). Panopticon is described in more detail in the previous section. VideoBoard was comprised of a static timeline on the left of a video player with each thumbnail representing every tenth second. The static thumbnails were animated on mouse-over events. This system was inspired by storyboard surrogates that are becoming increasingly popular [1] but added the feature of presenting all video frames on demand (i.e. mouse-over) similar to popular video editing programs like iMovie and Final Cut Pro. Finally, YouTube represents one of the most popular interfaces for video delivery [2] and has recently begun using preview thumbnails for context while seeking.

Three different videos (~30 min in length) were chosen to represent different application areas: an edited video, an unedited video, and a surveillance video. The edited footage – an extract from “Beverly Hills Ninja” – aimed to represent a standard film where users may want to find a specific scene (e.g. an action scene in a movie). The unedited footage – filmed sequences of a park, city center, and market area – aimed to represent personal videos and life-logging where users may want to find a specific event (e.g. when the family encountered a squirrel). Finally the surveillance footage – static camera footage of an office entrance during a normal working day – aimed to represent a pre-identified key time in a video where a user may want to find a specific person.

To assess gist comprehension, a forth video of the same duration was used – an extract from the movie “Soccer Mom”. We chose the movies (“Beverly Hills Ninja” and “Soccer Mom”) as they were freely available to download, had a chronological narrative, were in color, and were not known as bestselling titles.

Participants were asked four different questions for each video to assess how quickly specific occurrences could be found in the videos. The rationale behind the questions was the same for all videos: it was important that none of the occurrences were too close together, i.e., finding the answer to one question should not already give away the answer to the next one. Furthermore, it was important that participants understood the questions without knowing the videos. So questions never mentioned names or other things that would have been better understood if participants had seen the video before. Also, it was important that occurrences were distinct, e.g. there was only one tray being carried out of the room, and there was only one lorry turning in the road. We left out any technical terminology which may not have been understood by all participants.

Procedure

Once participants had been briefed, and informed consent collected, they were introduced to their randomly-selected system. Participants were given up to 10 minutes to explore and familiarize themselves with the different functionalities of their system, using both a practice video (“Big Buck Bunny”) and practice questions.

Following the practice, participants were given the four questions (in counterbalanced order) corresponding to the first video, both on paper and read aloud by the researcher. We made sure that participants understood the questions before starting a stopwatch to record completion time, and participants were asked to answer the questions verbally. The accuracy of the measured performance time was ensured through a video recording of the screen and additional audio recording of the overall session.

Once participants had satisfactorily answered all four questions for a video, the same procedure was used for the following two videos. The order of the videos presented was also counter balanced. To assess how well participants would understand the narrative of a video, a fourth video was presented at the end and participants were given exactly five minutes to review the footage. Afterwards, participants had to write a short summary about everything they had understood from the video.

Study 1: Quantitative Results

Search Times

A two-way mixed ANOVA was carried out on the time data with System (Panopticon, VideoBoard, YouTube) and Video (edited, unedited, surveillance) as the independent and repeated factors respectively.

A main effect of System was found, F(2,33)=20.567, p<.001, where participants were significantly faster with Panopticon (288.50 sec.) than with YouTube (1252.83 sec., p<.01) and VideoBoard (321.06 sec., p<.01). A significant difference was also found in speed between VideoBoard and YouTube (p<.05).

A main effect of Video was also present, F(2,32)=88.662, p<.001, where significant differences in search times were observed between all three video types (p<.001) with the unedited footage being the fastest, and surveillance video the longest.

An interaction effect was found between System and Video, F(4,66)=7.939, p<.001 (see Table 1) with no differences between systems for the edited video. However, for the unedited video YouTube was significantly slower than both Panopticon (t(22)=3.305, p<.010) and VideoBoard (t(22)=2.325, p<.05). No difference was found between Panopticon and VideoBoard.

Video	Panopticon	VideoBoard	YouTube	*F(df)*	*p*
Narrative movie	288.50 (85.40)	308.17 (134.21)	325.33 (163.37)	0.24 (2,33)	.79
Unedited footage	109.58 (31.47)	131.25 (66.01)	215.00 (105.92)	6.74 (2,33)	**< .01**
Surveillance footage	235.50 (117.48)	523.75 (195.91)	712.50 (124.16)	30.73 (2,33)	**< .01**
Total time	626.08 (179.12)	963.17 (303.56)	1252.83 (193.99)	21.88 (2,33)	**< .01**

Table 1. Mean times in seconds (standard deviations in brackets) for the four questions of each video (left) in each condition (top).

For the surveillance video, significant differences in search times were found between Panopticon and VideoBoard (t(22)=4.371, p<.001), Panopticon and YouTube (t(22)=9.667, p<.001), and between VideoBoard and YouTube (t(22)=2.819, p=.010).

Gist Comprehension

A list of requirements was created to evaluate the summaries written by participants of their understanding of the last video. This list contained 25 entries of the most important parts of the storyline (e.g. "Family: 3 kids"). Two researchers independently classified the summaries. Final scores were matched and emerging disagreements discussed and resolved. One point was awarded for each correctly mentioned entry.

System	Panopticon	VideoBoard	YouTube	*F(df)*	*p*
Mean Score (SD)	9.25 (2.96)	5.58 (2.88)	5.33 (2.87)	6.85 (2,33)	**<.01**
Range	5-14	1-10	2-11		

Table 2. Participants' scores in the gist comprehension task. ANOVA reveals significant differences between the conditions.

Table 2 shows the final scores of participants in each video condition. A one-way ANOVA found that participants included a significantly higher number of important aspects in their summaries when using Panopticon in contrast to the other video systems. (Panopticon vs. YouTube, p<.01; Panopticon vs. VideoBoard, p=.01; YouTube vs. VideoBoard, p>.05)

Study 1: Qualitative Findings

Following the completion of Study 1, participants were asked to fill out an open-ended questionnaire asking them about their experience using one of the three systems. These responses were analyzed using Content Analysis by one researcher and verified by a second researcher.

Participants were in agreement that the grid-style timeline – present in both Panopticon and VideoBoard – was an important feature. More specifically, participants agreed that the timeline allowed them to see an overview of the video and use that to scan the most appropriate areas of the video first.

Finding specific information was generally perceived to be more appropriate in unedited or surveillance footage rather than edited footage where fast transitions made it difficult for the viewer to focus. However, not all participants were in agreement, suggesting that the ease of spotting information depended on various factors.

A major difference between the two surrogates that participants pointed out was the absence of all frames from the overview in VideoBoard. Participants noted that while it was not an ideal implementation to have large a number of frames hidden, the timeline still helped in finding information faster than with a regular video player. The fact that the system allowed the participants to preview the hidden frames on demand was an added bonus.

Despite a generally favorable perception of Panopticon, participants pointed out a number of problems with the system. The most common issues included thumbnails that were too small and at times made it difficult to identify objects. Participants also commented that the inability to pause the timeline for a quick search made them feel as if they were being overloaded with information.

PANOPTICON REDESIGN

Study 1 was replicated using a Tobii X60 eye tracker with the aim of understanding participants' behavior when searching for information in videos. The variable under investigation comprised of Mean Visit Duration – the average time (in seconds) participant spent looking at a feature (timeline or video player) per visit to the area.

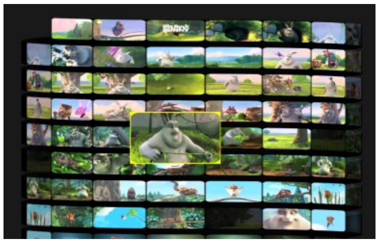

Figure 5. Panopticon 2 Interface, showing video playback within the overview.

It was observed that participants spent significantly longer on the timeline window than on the video window with Panopticon (p<.001) and VideoBoard (p=.001). These results support the qualitative findings indicating users' preference of using the timeline over the video window. It was also found that participants viewing the unedited (p<.010) and surveillance (p<.001) videos spent significantly longer on average on the timeline window than on the video window. Motivated by these findings, coupled with the qualitative feedback from Study 1, Panopticon was redesigned. This 'Panopticon 2' contained three core changes: the integration of the video player into the timeline, larger thumbnails, and a pause functionality.

Firstly, the video window was directly integrated into the timeline window (see Figure 5) in order to facilitate the search for information. In this configuration, the user may keep their focus on the timeline while spotting relevant events and objects in their peripheral vision [10]. Additionally, participants should not waste time switching between the timeline and video window – as indicated by the eye tracking gaze plots (see Figure 6 and Figure 7).

Secondly, the size of the thumbnails was increased as a result of participant feedback. For a given length of video, there is a trade-off between maximizing the thumbnail size (to retain resolution) and minimizing the interval between cells (so an overview is given in a short time). A practical compromise was reached by showing a windowed view over a range of the video, and the user may vertically scroll over the overview. This still maintains the coherency of the rendered output – exactly the same output appears at the same time in a loop, albeit at different vertical positions.

Finally, a key-press 'pause' functionality was implemented to satisfy participants who found the initial experience with the system overwhelming.

STUDY 2

A second study was carried out to evaluate Panopticon 2. Based on the changes made from participant feedback and the eye tracking results, it was expected that participants would be able to find information more quickly using this new system.

Method

A repeated measures design was used to evaluate the time taken (seconds) for participants to find information with Panopticon 2. Participants were required to find the information in the same videos: edited footage, unedited footage and surveillance footage. 13 new participants aged 18-24 (M=20.7 years old) took part in this study (8 female and 5 male). They were recruited from university and were compensated accordingly.

Participants were first taken through a demo with the new Panopticon version to familiarize themselves with it. They were then presented with the three videos and the related questions to each video. Videos and questions were both presented in a random order. Time was recorded for each question via stopwatch after the questions were read aloud.

Results

A two-way mixed ANOVA was carried out on the time data (seconds) of the three different videos and the video system (Panopticon, VideoBoard, YouTube, Panopticon 2) as between-subject factor. A main effect of video was found, F(2,44)=93.02, p<.001, with significant differences between all three videos (p<.001). A main effect of system was also present, F(3,45)=123.79, p<.001 showing significant differences between all systems (p<.010) with the exception of Panopticon and Panopticon 2.

An interaction was found between video and system, F(6,88)=17.12, p<.001, where participants searching within the surveillance video were significantly faster with Panopticon 2 (29 sec.) than with Panopticon (72.12 sec.), t(23)=3.131, p<.010. .No difference was found between systems for the edited video, but for the unedited video YouTube was significantly slower than both Panopticon (p<.010) and VideoBoard (p<.05). No Difference was found between Panopticon and VideoBoard.

DISCUSSION

The majority of past work on video surrogates has primarily focused on gist comprehension and largely ignored object recognition [3, 12]. In this paper, we have looked at object recognition as well as gist comprehension with Panopticon, a new video surrogate system. Results from two user studies demonstrated that participants using Panopticon were faster in recognizing objects than those using VideoBoard and YouTube with unedited footage (e.g. life-logging, personal video) and surveillance video. An improved version of the system, Panopticon 2, further enhanced the performance of users with surveillance footage, but not with the other two video types. Panopticon also proved to be significantly better than both VideoBoard and YouTube in gist comprehension.

Participants using the YouTube player were the slowest at finding information within videos and this could be attributed to the fact that they had to view more of the video due to less frames being available at once in the timeline. The eye tracking data found that participants using Panopticon and VideoBoard spent longer on average viewing the timeline, while no such difference was observed for participants using the YouTube player. In other words, participants using YouTube spent similar amounts of time viewing the timeline and watching the video. Given the differences in timeline design between the YouTube player and the other two systems, it is reasonable to suggest that YouTube's poorer results were due to the lack of a comprehensive overview.

Surveillance

Panopticon was clearly superior to the other two systems with the surveillance footage. This was the case across both studies and was further reinforced by participant feedback.

The heat maps from the eye tracking data (see Figure 6) clearly show how participants using Panopticon were able to identify the key areas within the timeline more easily than those using VideoBoard – given that very little time was spent looking away from the key areas. The fundamental difference in design between Panopticon and VideoBoard was the animation of the key-frames in the

Panopticon timeline. Since participants were significantly faster finding information in surveillance videos with Panopticon when compared with VideoBoard, it is possible to conclude that the animated key-frames were the main factor in the improvement of information search in surveillance. A look at the mouse clicks for the systems also shows that clicks were fairly equally distributed over the whole window for VideoBoard while they were concentrated in key areas for Panopticon. This observation is corroborated by feedback from participants who complained about the 'hidden frames' in the VideoBoard timeline, which resulted in them having to mouse over regions to find the required objects.

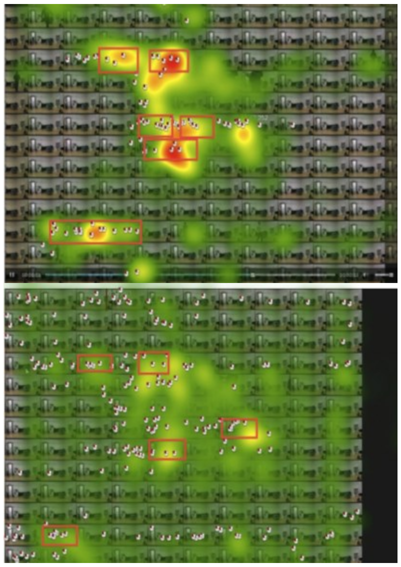

Figure 6. Panopticon (top) and VideoBoard (bottom) heat maps for surveillance video, with key areas (red boxes) and mouse clicks (red dots) shown.

Possible applications for Panopticon in the context of security could be twofold. In an implementation where an area is monitored automatically and alerts are triggered when unusual activity is detected, a generous amount of footage could be downloaded as a Panopticon movie in order to preserve all information from the footage while also saving browsing time and bandwidth. Instead of the offending event being presented to an expert in isolation, much richer contextual footage could be included. In an implementation where an area is monitored in real-time, Panopticon can be modified to work with a "live" view. For example, the lower-right corner showing "now", and "now-10" to the left of that, forming a grid of time. As the "now" view is constantly updated, the other thumbnails would not have to move in order to keep this relative, yet consistent, spatiotemporal layout. This view could serve as a fall back system where the user can quickly skim footage from the past few minutes for missed information. In both of the described scenarios Panopticon could add value to surveillance systems, and the usability of such a system amongst security professionals should be evaluated.

Edited Video

Interestingly, the YouTube player was comparable to VideoBoard and Panopticon for finding information in an edited video. A look at the heat maps for the edited video for Panopticon and VideoBoard (see Figure 7) shows little difference in spread of gaze time for both systems, suggesting that participants using Panopticon were not able to identify the key areas of the video as easily as with, for example, the surveillance video (Figure 6). As a result, participants had to employ a seeking approach similar to that used with VideoBoard. Participants viewing the edited video were required to view more of the video in the video player than participants viewing the other two videos – as demonstrated by the Feature and Video interaction – further supporting the notion that participants were unable to quickly spot important segments in the timeline.

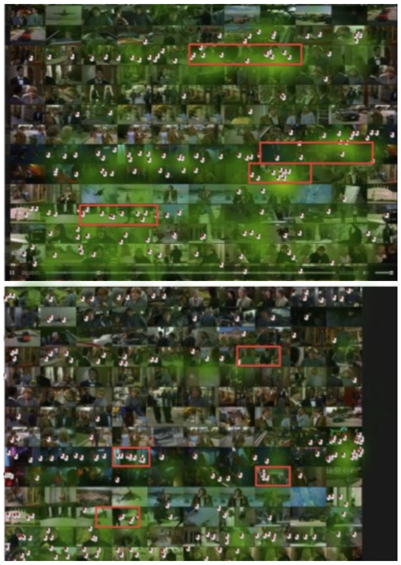

Figure 7. Panopticon (top) and VideoBoard (bottom) heat maps for edited video, with key areas (red boxes) and mouse clicks (red dots) shown.

A potential interpretation of this finding, supported by participant feedback, is that the storyboard timeline design does not lend itself for searching in edited, narrative-driven videos with fast transitions. Participants had to resort to watching more of the video which meant that they were using the surrogate system in a similar way to the video browsing system – and hence no significant differences in search times. Previous work seems to suggest that repetitive footage is not represented well with surrogates [14] but this notion was not supported by our findings.

Limitations

There are a number of limitations that have to be taken into consideration in this study. The questions used for the surveillance video were slightly different to those used for the other two videos. Two questions for the surveillance footage asked participants to count actions (e.g. the number of times a person passed through a door) which added another layer to solely searching. The different questions explain why the surveillance video took the longest to complete across all systems, but as the same four questions were used for all three systems the differences found should be viewed as legitimate. Another limitation was that only one video was used for each type. With this in mind, it is possible that the video that was chosen for the edited condition happened to not suit storyboard timelines due to its structure and fast action sequences. Finally, this study evaluated the performance of participants searching for information in previously unseen videos. However, a common use case is to search in videos users are already familiar with, and it would be interesting to look at this scenario. For example, video editing with footage a user has shot themselves, or a user finding a scene in a movie they just finished watching in order to clarify plot details.

CONCLUSION

This paper presents a new video surrogate system, Panopticon, which does not remove any frames and uses a novel animated arrangement to maximize context to improve gist comprehension, object and event recognition in videos. The overview can be generated as a highly efficient repeated animation loop for resource-constrained environments. As an example, a 30 minute video can be presented as a 10 second loop which allows any sub-sequence (even the full 30 minutes) to be seamlessly followed.

A series of user studies demonstrated the superiority of the system with both unedited footage and surveillance footage over a popular video browsing system – YouTube. In addition, Panopticon was shown to be superior to another video surrogate system with surveillance footage, suggesting a particular niche in that application area. The eye tracking studies suggest that the advantage is the result of the animated timeline.

Much work remains in fully exploring this new technique. These studies have only used a uniform grid layout, and more general transform function need exploring (e.g. larger, slower moving cells where they represent more salient scenes). Additionally, despite the multimodal nature of video, additional data such as speech, music, and subtitles are not yet represented within the Panopticon.

ACKNOWLEDGEMENTS

We would like to thank Dermot McNaney, Tom Bartindale and Robyn Taylor for work on the associated video. This work was partly funded by the SiDE project.

REFERENCES

1. Christel, M. (2008). Supporting video library exploratory search: when storyboards are not enough. *In Proc. CIVR 08* (pp. 447–456).
2. Cunningham, S., & Nichols, D. (2008). How people find videos. *In Proc. JCDL '08* (pp. 201–210).
3. Ding, W. and Marchionini, G. A Study on Video Browsing Strategies. Tech. Report CS-TR-3790, UMIACS-TR-97-40, CLIS-TR-97-06, University of Maryland. 1997.
4. Divakaran, A., & Otsuka, I. (2007). A video-browsing-enhanced personal video recorder. *In Proc. ICIAPW '07* (pp. 137–142).
5. Gil, N., Silva, N., Dias, E., & Martins, P. (2012). Going through the clouds: search overviews and browsing of movies. *In Proc. MindTrek 12* (pp. 158–165).
6. Haering, N., Venetianer, P. L., & Lipton, A. (2008). The evolution of video surveillance: an overview. *Machine Vision and Applications*, *19*(5-6), 279–290.
7. Hollingworth, A. (2009). Two forms of scene memory guide visual search: Memory for scene context and memory for the binding of target object to scene location. *Visual Cognition*, *17*(1-2), 37–41.
8. Marchionini, G., Song, Y., & Farrell, R. (2009). Multimedia surrogates for video gisting: Toward combining spoken words and imagery. *Information Processing & Management*, *45*(6), 615–630.
9. Matejka, J., Grossman, T., & Fitzmaurice, G. (2013). Swifter: Improved Online Video Scrubbing. *In Proc. CHI 2013*. pp. 1159-1168.
10. Palmer, S. M., & Rosa, M. G. P. (2006). A distinct anatomical network of cortical areas for analysis of motion in far peripheral vision. *The European Journal of Neuroscience*, *24*(8), 2389–405.
11. Schoeffmann, K., Hopfgartner, F., Marques, O., Boeszoermenyi, L., & Jose, J. M. (2010). Video browsing interfaces and applications: a review. *SPIE Reviews*, *1*.
12. Song, Y., & Marchionini, G. (2007). Effects of Audio and Visual Surrogates for Making Sense of Digital Video. *In Proc. CHI 2007* (pp. 867–876).
13. Tse, T., Marchionini, G., Ding, W., Slaughter, L., & Komlodi, A. (1998). Dynamic key frame presentation techniques for augmenting video browsing. *In Proc. AVI 98* (pp. 185–194).
14. Wildemuth, B., Marchionini, G., Wilkens, T., & Yang, M. (2002). Alternative surrogates for video objects in a digital library: users' perspectives on their relative usability. *LNCS*, *2458*, 493–507.
15. Wildemuth, B., Sun Oh, J., & Marchionini, G. (2010). Tactics used when searching for digital video. *In Proc. of IIiX '10* (pp. 255–264).
16. The Zettabyte Era. Cisco Visual Networking Index White Paper, 2012. http://cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/VNI_Hyperconnectivity_WP.pdf