Visualization to Aid Video Navigation

A Proposal for the Second Year Project in Cognitive Science

Adam Fouse

October 4, 2008

DRAFT 1

ABSTRACT

The ready availability of digital video is revolutionizing the way that observational science research is performed. Digital video has made it easy to record, store, and share vast amounts of video data about human activity. However, this vast increase in the availability of video has not yet been accompanied by equally advanced mechanisms for understanding and analyzing the content of these videos. I propose a project to develop better methods of visualizing the content of videos by first understanding the cognitive issues at play when people view and analyze video. These include questions about episodic memory for video, summarization and segmentation of video data, and perception of video across a range of auditory and visual characteristics and playback speeds. This project will require three stages: First, I will conduct an experiment on video navigation to analyze the strategies that people use to navigate video, which guide further inference about the cognitive factors at play. Second, I will design visualizations to take advantage of strengths and support weaknesses in those factors. Third, I will use similar methods as in the first experiment to evaluate the design of the visualizations against the identified cognitive factors.

1. Introduction

The ready availability of digital video is revolutionizing the way that observational science research is performed. Digital video has made it easy to record, store, and share data about human activity. Inexpensive recording and storage mechanisms have allowed researchers to accrue vast amounts of data concerning human activity. Many research sites are instrumented with multiple video cameras, leading to situations where single recording sessions can result in several hours of data. This vast amount of data represents incredible opportunities for analysis of human activity to gain insight into in-situ human cognition. Much of this video is of activity in everyday living and working environments, and even video of closely controlled experiments provides insight into subtleties of behavior that may be missed by traditional response mechanisms. Analysis of these sources of data may lead to crucial insights into human behavior and cognition.

However, this vast increase in the availability of video has not yet been accompanied by equally advanced mechanisms for understanding and analyzing the content of these videos. Existing methods for analyzing video frequently use basic personal computer media players, with controls based on those of a VCR. The controls for VCRs were limited by the linear nature of videotape. Digital video does not have the same constraints, but this lack of linear constraint has not been fully leveraged to design better analysis tools.

One promising way to take advantage of this lack of linearity is to design tools that utilize advanced visualization of the content of videos. Information visualization is an active research field, and has been acknowledged by many researchers to be a promising method of making information easier and quicker to understand. Technological advances have made the creation of compelling interactive visual representations of complex information possible on everyday personal computers. However, little attention has been paid to visualizing he content of videos. Better ways of visualizing the content of video are needed to allow for more efficient understanding and navigation of video. A good visualization of video might allow a viewer to understand the general content of a video at a glance, and to quickly find the particular segment that might be of interest.

In addition, while much effort has been put toward the development of new visualization methods, less effort has been put toward the evaluation of these efforts. With any piece of technology, it is important to have an empirical understanding of whether it does any good. When evaluation is performed for visualization efforts, it frequently does not provide deep understanding into the factors that influence why it works or doesn’t work, and what situations it under which it works better. Useful evaluation needs to consider how the visual representation of the information is perceived, how it will be used by the task, and what contextual factors will affect its interpretation. Techniques for this type of evaluation exist individually, but rarely are combined for the comprehensive evaluation of visualization that is needed.

Performing empirical evaluation of visualization requires developing some understanding of the user’s perception and comprehension of the information that is being presented. In this case, developing better visualization methods for video requires developing a better understanding of the cognitive issues that are at play when people interact with video. A range of related cognitive research both guides this effort as well as identifies questions that need to be answered.

An important set of questions is centered on the representation of video sequences in memory. What are the semantic and perceptual aspects of video data that are remembered? There is much known about memory for pictures, as well as personal episodic memory. Does memory for video data utilize the same mechanisms? If so, what qualities of episodic memory might impact the recognition of important segments of video? What characteristics of human memory might affect the ability to remember important segments that have already been viewed, and what does this suggest about how videos should be presented and analyzed? How do people identify important segments, and what elements of a video recording are used people summarize the content? How can this information guide automated processes to condense and summarize?

Another important question regards the perception of video data. How can existing knowledge about perception of motion be used to maximize the perception of important segments of video with and without significant visual motion? An important aspect of digital video is that it can be viewed at a variety of positive and negative speeds, rather than just at real time. There has been some research regarding perception of sped up recording of speech, but there are many questions that remain that may provide some insight into how video analysis tools should be designed. For example, how fast can people reliably perceive the content of an audio and video recording with the goal of identifying key segment of the video? What aspects of the data will be most often overlooked?

A third consideration is the differences between different types of video. Researchers will often work with video from different angles, and first-person video from a natural angle and height may be perceived differently than a video shot from overhead or through a fish-eye lens. Alternatively, the video may not be through a video camera at all, but rather a screen recording from a personal computer. Screen recordings may utilize different perceptual abilities than video of humans.

The above listing of cognitive and perceptual issues involved in video analysis is not comprehensive, but rather indicative of the difficult task of developing better tools for video analysis. Investigation is needed to study the complexities and difficulties that are involved in video analysis, and also how current tools may shape some of the cognitive issues involved. If we hold the view that cognition is the product of the brain, the body, and the external environment, then it follows that current tools may affect the answers to the above questions. A better understanding of this interaction should lead to the design of better tools.

There are many aspects of working with video that can be investigated. A user needs to be able to understand the content of video and navigate to a target location. For scientific uses, analysis and annotation of video need to be supported in ways that make annotations easier to create and to share. With large video libraries, better ways need to be developed to understand the information that is spread out over multiple videos, and to compare the content of these videos. However, for the purposes of this proposal, the focus will need to be narrowed from this broad collection.

Therefore, I propose a project to investigate video navigation, with three general goals. First, to understand the cognitive issues involved in navigating digital video, with the hope of developing an understanding of what human, data, and tool characteristics affect video navigation tasks, and where existing interfaces fail to support these tasks. Second, use these results to design and develop new ways to visually represent the content of videos. Third, empirically evaluate the new visualization methods in the context of the perceptual and cognitive issues identified under the first goal. It is hoped that this process will serve a direct goal of improving video navigation and a secondary goal of informing future empirical visualization work.

2. Background

**** Note: This section is a work-in-progress. I need to spend a little more quality time with the literature, and is currently centered on research that came out of the HCI domain. Things I plan to improve include:*

- *Cognitive and perceptual issues background material*
- *Expand and refine descriptions of existing related work to target a broader audience*
- *Include relevant images and figures*

In the domain of video navigation, there have been several research efforts to understand consumer video navigation in the context of traditional VCR-style controls. These studies have included some empirical evaluation, particularly as it relates to navigating a collection of video clip in search of target information (Crockford and Agius 2006). Other studies have looked at proper level of skipping forward (Drucker, et al. 2002), speeding up playback for browsing content (Furini 2008), and simultaneous playback of multiple speeds of video (Eerenberg, Aarts and de With 2007). In addition, there are recent efforts focused on the indexing and searching of video as an aid to navigation (Smeaton 2007). This work looks to move beyond the linear model of tape video and support new means of interacting with digital video.

The previous studies are largely focused what control abilities should be provided for video navigation, but another area of research focuses on ways to represent video to make it easier to quickly understand the content of video. This research can be broadly divided into two general areas: automatic summarization and novel visual representations.

A growing number of techniques have been developed to aid in automatic video summarization (Li, Zhang and Tretter 2001). Computer vision techniques, such as computing visual angular velocity (McCall, et al. 2004) or spatio-temporal analysis (Xiao, et al. 2008) have frequently been used to provide advanced automatic summarization. Other efforts take a more semantic approach, looking to extract meaningful sub-units from a video (Yeung, Yeo and Liu 1996) or create an understanding of the underlying structure (Pope, et al. 1998).

Visual representations of video content have also been created. These tend to rely on human abilities to make sense over a condensed version of the entire video, rather than selecting specific points in the video. The Recreating Movement project (Hilpoltsteiner 2005) demonstrates a technique of stacking frames on top of each other in three-dimensional space. This technique is particularly applicable to the analysis of physical movements. In videos with a single human subject, a chroma-key is used to isolate the subject from the background, and can reveal a range of movement at a single glance much better than with a traditional still frame or set of frames. A more abstract technique for visually representing the content of a video is to use slices of visual information from each frame (Nunes, et al. 2006). In this case, the visualization uses a single column of pixels near the center of each frame that are then combined to create an abstract timeline representation of the video. The contiguity and visual flow of colors provides an indication of areas of activity and change, providing a high-level overview of the data and aiding further analysis.

There is a small but growing acknowledgement of the importance of perceptual evaluation of information visualization methods (House, Bair and Ware 2006, Ware 2000). These methods are often based in the application of known perceptual properties to the design of visualization. One characteristic that has often been evaluated is the accuracy with which information is transferred. For example, a color scale that is linear in numeric value may not be perceived as linear in the information it represents (Ware 1988). Further, the chosen scale for representing information may be too fine grained to provide benefit for information perception, and may simply add clutter without improving perception of information (Bisantz, Marsiglio and Munch 2005).

3. Methods

****Note: The general framework for this project is accurate, but some of the details need further refinement. In particular, I haven't yet decided upon the right experimental task. In this section, I need to do come to a conclusion with regard to the task, and then do a better job of explaining why this is the right way to go.*

The general goal of this research path is to understand the perceptual and cognitive issues related to analyzing digital video. As described in the introduction, there are both a wide range of issues and wide range of task types to be considered. In general, the cognitive issues seem to cut across the task types; most issues will affect most tasks, but in different ways. As such, given the scope of this proposed project, I plan to restrict my investigation to a single type of task with the hope of beginning to understand a range of the cognitive issues at play, which can then be further investigated with future efforts. Video navigation is a natural starting point. Other tasks that could be considered include human and automatic summarization, selecting of individual clips from a library, annotation, and analysis of annotation. All of these tasks require navigation of a video clip to complete the task. A better understanding of the complexities of video navigation should provide a base on which to build understanding of the other tasks, and guidance for fruitful experimental directions.

This project will have three components. First, I plan to conduct experiments to investigate video navigation performance with common existing user interfaces. Second, I plan to use the results of this experiment to design visualizations of video context for the specific goal of improving navigation. Third, I plan to empirically evaluate these new visualizations to determine whether they are able to help people perceive and understand the content of videos in a more efficient way. In this proposal, the most detail is given about the first component, since the following components will be partially defined by the results of the first set of experiments.

3.1. Experiments with current interfaces

The first set of experiments will test video navigation performance with three existing interfaces. These will include VCR-style controls, computer media player style controls, and a filmstrip interface (*** A figure will be included here with examples of each interface). The VCR-style controls will include buttons to play, pause, fast-forward, and rewind. The media player style controls will include those of the VCR controls, with the addition of a draggable position slider to indicate current time point in the video as well to allow the user to set the current time point (the video display will be updated while the user is dragging). The filmstrip controls will include those of the media player, with the addition of a "filmstrip" visualization that will display frames from the video clip, chosen at equal intervals such as to fill a horizontal bar.

There are two subject populations that I plan to include in these experiments. The first population will be naïve subjects, such as undergraduates. It is assumed that most people will have had experience viewing digital video and be familiar with the major concepts, but this assumption will be screened through pre-experiment questions. These will follow through a carefully controlled experiment using ethnographic video of human activity. The specific videos chosen will depend on availability, but the goal will be to select videos that are have elements for analysis that are easily understood by a naïve subject but still require directed attention to identify. The videos will need to share the characteristic of having visually distinguishable time points for which to search. They will view the videos through custom Flash video software that replicates existing interface types and will maintain consistent design between the different interfaces when possible.

The second population will be experienced researchers. These will be people that analyze video in the course of their work, such as those using observational data to make conclusions about human or animal behavior. I hope to have these subjects participate both in the controlled experiment and also in observation of their normal video analysis methods. For the observational side of the study, I hope to collect information about researchers analyzing familiar data with familiar tools. The goal will be to collect information about the ways they interact with video data in the course of their research.

The task for the controlled experiment will need to one that properly engages the participants in ecologically-valid activity and requires understanding of the high level structure of the video as well as low level details. It should also not require expert level knowledge of either the domain or the tools; it should engage the participant in understanding the video rather than struggling with the task. As such, there are two tasks that are currently under consideration. Pilot experiments and further thought and collaboration will be used to select and refine one of these tasks.

The first task under consideration would be to analyze a video by looking for instances of a group of specific actions in a video clip, such as specific types of hand gestures. They will then have to view the entire video, then go back and mark all of the time points at which this action occurs. This task will require the subjects to navigate the entire video while engaging in analysis of the content of the video. The accuracy and time to completion will be measured and compared across trials.

The second task under consideration would be to analyze the video by dividing it into meaningful segments. In this case, participants would view the entire video, then go back and mark the segments of the video and select which segments are meaningful to the content of the video. This task would also require subject to navigate the entire video, but would shift the focus more toward a high level understanding of the structure of the video. The navigation required would be less predictable and directed, as when compared to the first task, which would have both positive and negative implications for the goal of understanding video navigation. It would provide more variability in the data, which should improve the ability to infer cognitive strategies from the results. Rather than searching for specific instances of actions in the video, participants will be searching for less strictly defined segment boundaries. This may result in a different type of navigation behavior, but might provide more insight into general video interaction and provide interesting information to guide future efforts in automatic summarization.

Regardless of the chosen task, screen recordings as well as overhead video recordings will be made. Each participant will be tested twice on each interface type, with a different video for each interface type. An assumption will be made that the participants are not familiar with the content of the videos; during the first trial for each video, the participants may need to get a good overview before searching for the specific point, and this should be lessened for the second trial. The two-viewing approach is meant to exert some control over the effect of familiarity with the video.

The results from this first pair of experiments will be used in several ways. First, there will be a quantitative comparison of the three interfaces that are tested, mostly focused on the results from the first subject population. This should be useful, both as a direct evaluation of those interfaces, and also to derive insight from those comparisons. While these results should show which interfaces work best in different situations, a comparison of the magnitude of difference between the various interfaces will help identify the cognitive issues that are affecting performance of the task and help with inference about why this might be the case.

Second, observational data in the form of screen recordings and video will be collected so that navigational strategies can be analyzed to further gain insight about the cognitive issues involved in video navigation. Data from both subject populations will be used, so that a range of strategies can be identified. It is hoped that particular difficulties or inefficiencies can be identified from this data.

3.2. Design of video navigation aids

The next element of this proposal is to use the results of the experiment to guide the design of visual representations to support video navigation. I will explore a range of visualization possibilities, focusing on representations that will address the difficulties and inefficiencies that are identified from the experimental and observational data. I plan to explore several possibilities, and use an iterative design process and informal testing to select and refine visualization methods that may be helpful.

3.3. Evaluation of video navigation aids

Once a set of video navigation aids is adequately refined, the final element of this proposal will be to empirically evaluate these visualization methods. The evaluation will focus on understanding the way people understand the information that is visually presented, especially in the context of the specific elements of video navigation that have been identified as difficult. The video search task from the first experiment will likely be repeated with the new visualization.

Depending on the visualization methods that are designed and the tasks that are identified, this evaluation may focus on specific elements of the visualization and require additional evaluation methods. For example, two alternative methods for displaying a condensed representation of segments of video may be developed, and the evaluation may focus on a comparison of these methods. In this case, additional tasks may be used in addition to the original search task, such as requiring subjects to summarize segments or to compare altered segments for differences.

4. Possible Results

The hoped-for results of this project are a better understanding of the cognitive issues that affect video navigation; better visualization methods for aiding in video navigation; and the beginning of an understanding of how people work with video in general, which can be used for future developments. This project has partially been designed with the idea that it will lead to future research. As mentioned in the introduction of this proposal, there are numerous areas within the area of visualizing video information, some of which may naturally continue from this line of research.

Specific to the individual components of this research proposal, the results from the first experiment will first provide a quantitative comparison of the three interface types. The expected result is that the interfaces that provide more information and interaction possibility will perform better than those that provide less. However, it is possible that this may not be the case, and that the additional information provided in the interface may be more distraction than help, and degrade navigation performance. It is also quite possible that there may be some subtleties in the results, where a simpler interface may perform better for some aspect of the task but the more complex interface performs better overall. This would be a best-case scenario, because results like that would help to create a description of how people perform the task.

From the observational data, I hope the results will allow for a description of the activity of navigating digital video, and a possible delineation of the strategies types of specific actions that are used in the process. The comparison of naïve participants to experienced video analysts may prove to be illuminating in a number of ways. Comparisons will be made with respect to level of experience working with video, different strategies based on the goals of the task (i.e., larger analysis goals versus immediate experimental goals), and familiarity with the videos. It is also possible that the observational data will not yield significant insight into the task of video navigation, in which case the planned second half of this project will have to be rethought and adjusted.

The results from the design and evaluation of new visualizations for aiding video navigation are hoped to be visualizations that have been shown to help users understand the content of video.

A final result that I hope to glean from this project is a better understanding of the process of designing and evaluating visualizations in the context of cognitive science. As was mentioned in the introduction, most visualization research has been focused on how information *can* be displayed. This is an important first step in the visualization design process, but I want to focus on how information *should* be displayed. This attempt to bring more science into the existing design and engineering of visualization is somewhat of a pilot of methods that I hope can be refined over a longer span of time.

5. References

Bisantz, A. M., S. Marsiglio, and J. Munch. "A comparison of the effects of graphical, linguistic, and numeric representaitons of uncertainty on decision-making." *Human Factors*, 2005.

Crockford, C, and H Agius. "An empirical investigation into user navigation of digital video using the VCR-like control set." *International Journal of Human-Computer Studies*, Jan 2006.

Drucker, S, A Glatzer, S De Mar, and C Wong. "SmartSkip: consumer level browsing and skipping of digital video content." *Proceedings of the SIGCHI conference on Human factors in Computing Systems*, Jan 2002.

Eerenberg, O, R Aarts, and P de With. "System Design of Advanced Video Navigation Reinforced with Audible Sound in Personal Video Recording." *Consumer Electronics, 2008. ICCE 2008. Digest of Technical Papers. International Conference on,* Dec 2007: 1 - 2.

Furini, M. "Fast play: a novel feature for digital consumer video devices." *Consumer Electronics* 52, no. 2 (Jan 2008): 513-520.

Hilpoltsteiner, M. *Recreating Movement.* 2005. http://www.recreating-movement.com/ (accessed 6 12, 2008).

House, D, A Bair, and C Ware. "An approach to the perceptual optimization of complex visualizations." *IEEE Transactions on Visualization and Computer Graphics* 12, no. 4 (Jul 2006): 509 - 521.

Li, Y, T Zhang, and D Tretter. *An overview of video abstraction techniques.* Technical Report, Palo Alto: HP Laboratories, 2001.

McCall, J, et al. "A collaborative approach for human-centered driver assistance systems." *IEEE Conference on Intelligent Transportations Systems.* 2004.

Minerva Yeung, Boon-Lock Yeo, and Bede Liu;. "Extracting story units from long programs for video browsing and navigation." *Multimedia Computing and Systems, 1996., Proceedings of the Third IEEE International Conference on*, May 1996: 296 - 305.

Nunes, M, S Greenberg, S Carpendale, and C Gutwin. "Timeline: Video Traces for Awareness." *Video Proc. ACM CSCW*, Jan 2006.

Pope, A, R Kumar, H Sawhney, and C Wan. "Video abstraction: summarizing video content for retrieval andvisualization." *Signals*, Jan 1998.

Smeaton, A. "Techniques used and open challenges to the analysis, indexing and retrieval of digital video." *Information Systems*, Jan 2007.

Ware, C. "Color sequences for univariate maps: theory, experiments and principles." *Computer Graphics and Applications, IEEE* 8, no. 5 (Sep 1988): 41 - 49.

Ware, C. *Information visualization: Perception for design.* New York: Morgan-Kauffman, 2000.

Xiao, Ruo-gui, Yan-yun Wang, Hong Pan, and Fei Wu. "Automatic Video Summarization by Spatio-temporal Analysis and Non-trivial Repeating Pattern Detection." *Image and Signal Processing, 2008. CISP '08. Congress on* 4 (Apr 2008): 555 - 559.