DRAGON: A Direct Manipulation Interface for Frame-Accurate In-Scene Video Navigation

Thorsten Karrer[†] **Malte Weiss**[†] **Eric Lee**[‡] **Jan Borchers**[†]

[†]Media Computing Group
RWTH Aachen University, Germany
{karrer, weiss, borchers}@cs.rwth-aachen.de

[‡]Apple Inc.
2 Infinite Loop, Cupertino, CA 95014
el@apple.com

ABSTRACT

We present DRAGON, a direct manipulation interaction technique for frame-accurate navigation in video scenes. This technique benefits tasks such as professional and amateur video editing, review of sports footage, and forensic analysis of video scenes. By directly dragging objects in the scene along their movement trajectory, DRAGON enables users to quickly and precisely navigate to a specific point in the video timeline where an object of interest is in a desired location. Examples include the specific frame where a sprinter crosses the finish line, or where a car passes a traffic light. Through a user study, we show that DRAGON significantly reduces task completion time for in-scene navigation tasks by an average of 19–42% compared to a standard timeline slider. Qualitative feedback from users is also positive, with multiple users indicating that the DRAGON interaction felt more natural than the traditional slider for in-scene navigation.

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation]: User Interfaces; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems.

General Terms

Algorithms, Design, Human Factors, Performance.

Author Keywords

Video Interfaces, In-Scene Video Navigation, Direct Manipulation, Object Dragging, Interaction Techniques.

INTRODUCTION

With the continued increase of the creation and consumption of continuous time-based media such as audio and video, better interaction techniques to navigate and manipulate these media are required. Today, even the average home user has the capability to produce, distribute, and consume digital video from sources such as YouTube (http://youtube.com), without expensive hardware, software, or training. As a result, there is an increasing need to support not only professional video editors, but also casual home users in video navigation and editing tasks.

One such task is frame-accurate navigation through a video scene. Frame-accurate browsing is not only important for video creation and consumption, but it is also a common task when, for example, analyzing clips of live sports events, annotating video, or reviewing recordings of scientific experiments. In these examples, the user typically focuses on a small number of interacting objects in the scene. This spatial interaction of objects is what gives a time point in a video its semantic meaning. Current interaction techniques for video navigation use time only as a syntactic construct. The linear timeline sliders found in most media players, fisheye-style warped timelines [6] and dynamic zoom sliders [2] allow the user to move through the timeline of the video; however, they all have in common the drawback that there is no direct relationship between a user's gestures and the actual content of the video. As a result, accurately pinpointing a particular moment in a video clip can be difficult.

Video navigation encompasses a large range of tasks, including coarse-grained navigation such as searching for chapters, and fine-grained (or in-scene) navigation such as scrolling to a frame where the objects in the scene are in a certain arrangement. In this paper, we focus on the latter scenario, where frame-accurate temporal pinpointing is required. We first present DRAGON (**DRAG**gable **O**bject **N**avigation), a technique for navigating video scenes based on the spatio-temporal evolution of the video content. This technique implements Shneiderman's notion of direct manipulation [8], allowing users to navigate through the scene by directly dragging objects to a desired location. We then present the results of a controlled experiment and user evaluation comparing our object-dragging technique to the standard timeline slider present in most video navigation interfaces today.

RELATED WORK

Previous attempts to improve the efficiency and accuracy of navigating through video address the problem more generally. Ramos and Balakrishnan's work on the *PVSlider* and the *TLSlider* focuses on pen-based interfaces to augment the timeline slider for video navigation [6]. Hürst's work on the *Zoomslider*, *NLSlider*, and *Elastic skimming* [3] for navigat-

CHI 2008, April 5 - 10, 2008, Florence, Italy.

ing through video addresses mapping problems between the video timeline and the slider widget itself. While all of these works improve upon the timeline slider, DRAGON leaves the slider metaphor entirely, replacing it with direct object manipulation in the video itself.

In [9], Zhai et al. discuss the drawbacks of the scrollbar widget for fine-grained positioning tasks in text documents; in [5], Lee discusses the similarities between the scrollbar for text navigation and the timeline slider for temporal (audio and video) navigation. Their findings imply that alternative approaches to video navigation should be investigated.

Very recently, Kimber et al. proposed *Trailblazing* [4], a system for multi-camera video surveillance that employs a similar interaction technique based on object trajectories. Their system, however, relies on object recognition and tracking to create the movement trajectories whereas DRAGON uses optical flow so that even individual parts of objects (like the hand of a soccer player) can be dragged around for navigation.

INTERACTION

Shneiderman defines a *direct manipulation* interface as one with "visible objects and actions of interest, with rapid, reversible, incremental actions and feedback" [7]. DRAGON allows users to directly manipulate the video contents by clicking on an object of interest and dragging it through the scene towards its desired position (and thus, time).[1] Dragging the object causes the video to scroll through time so that the object follows the mouse pointer (see Fig. 1). Note that the user does not have to stay exactly within the object's trajectory through the video: as with most draggable UI elements (e.g., common scrollbars) the object behaves as if it were on rails, but attached to the pointer with a rubber band.

For example, consider the video of a car approaching an intersection like the one shown in Fig. 1. As the car approaches the intersection, it slows down, yields to a passing car, and turns at the intersection, accelerating as it leaves the scene. The user wishes to navigate to a point where the car occupies an interesting position to check, for example, if the traffic light is still green when the car entered the intersection. This positioning task can be difficult using a slider-like control, as the slider is mapped linearly to absolute time, but the car's acceleration and deceleration result in a non-linear position-time relationship. With DRAGON, the user can click on the car when it enters the frame and drag it along its trajectory through the intersection to its desired position.

This direct manipulation technique overcomes two drawbacks of the timeline slider:

- The direction of the user's gesture is directly related to the direction that the object of interest moves in the scene. In contrast, dragging a timeline slider from left to right can result in an object in the video moving arbitrarily, even from right to left, depending on the contents of the video.

[1]We invite readers to view the video figure accompanying this submission for a demonstration of the DRAGON interaction technique.

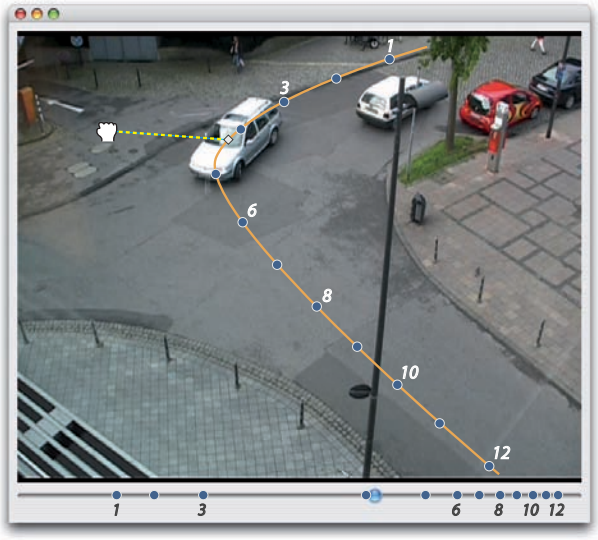

Figure 1. The DRAGON interaction technique. The user has clicked on the car at the diamond marker, and is now free to drag the car along its trajectory in the video. As she moves the car, the video scrolls through time accordingly. As the car is attached to the mouse pointer with a "rubber band", the user is not required to stay exactly on the car's movement trajectory in the video. Notice how evenly spaced positions on the car's trajectory correspond to a non-linear temporal spacing on the timeline slider.

- The size of the user's gesture is directly correlated to how far the object of interest moves in the scene. In contrast, dragging a timeline slider one pixel can result in the object moving several pixels, or even none at all.

IMPLEMENTATION

To obtain the object trajectories that DRAGON uses to support direct object manipulation, we use an algorithm based on work introduced by Brox et al. to precompute *optical flow* fields between neighboring frames in the scene [1]. The optical flow fields provide an estimate of where each pixel in a video frame moves to in the succeeding frame, and where it came from in the preceding frame (see Fig. 2).

Calculating the optical flow fields for each second of standard definition video currently requires over 15 minutes of processing time on a quad-core 3 GHz Mac Pro. However, this processing can be performed offline, and the results are stored together with the video on disk. Real-time interaction is achieved at runtime by using these precomputed flow fields to calculate the required object trajectories of the pixel the user clicks on. When an object is dragged, it follows its trajectory in both space and time. That is, we look for the frame where the object's three-dimensional (x, y, t) distance to the mouse pointer is minimal (see Fig. 3). This ensures that there are no large jumps in the video playhead position, and also allows us to correctly disambiguate situations where an object appears at the same spatial position in the video at multiple points in time (e.g., a video of rotating hands on a clock). Further details about the implementation of DRAGON can be found on the web[2].

[2]http://hci.rwth-aachen.de/dragon

Figure 2. The flow field (shown in white) stores the pixel correspondences between pairs of adjacent frames. These correspondences allow us to recreate the movement trajectories of arbitrary objects in a scene, such as the billiard balls in this table shot.

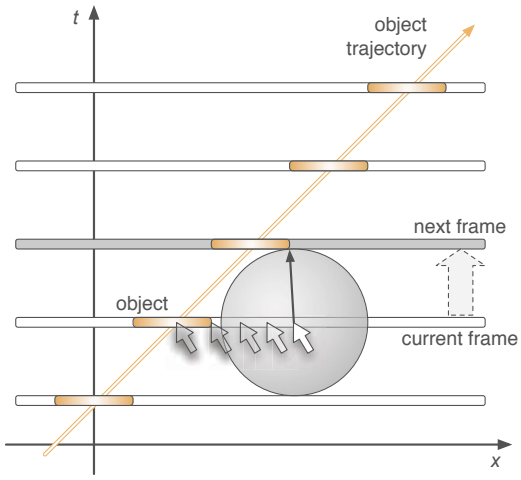

Figure 3. Top view of a stack of frames (i.e., the y-axis is pointing out of the picture plane towards the reader). When the user clicks on the object and moves the mouse to the right, the video is scrolled to the frame where the (x, y, t) distance between mouse pointer and object is minimal. This distance is measured in both space and time, represented in the diagram by the shaded sphere, to avoid unwanted jumps along the video timeline.

EVALUATION

We hypothesized that DRAGON would be significantly faster than the traditional timeline slider for frame-accurate in-scene navigation tasks, and designed a user study to compare task performance times between these two navigation techniques.

Experimental Setup

30 participants (21 male, 9 female) between the ages of 22 and 39 who use computers regularly were recruited to participate in the study. Participants were asked to perform navigation tasks on four different billiard scenes. At the start of each session, we allowed the users to familiarize themselves with both DRAGON and the timeline slider for navigating through a test video; most users were already familiar with the timeline slider, as it is a standard control on all software multimedia players. After that, users were asked to perform a specific navigation task on each of the four scenes using both DRAGON and the timeline slider (yielding a total of eight measurements per user). For each scene, users were first shown the video once, before being given instructions on the navigation task to be completed for that video. Then, they were shown the video once again, before finally being asked to perform the navigation task. As each user was exposed to all data sets, we were able to stagger the order in which they were presented, and counter-balance the order of the technique that was used to complete the task, to minimize learning effects. Each session lasted approximately 15 minutes, and concluded with a brief questionnaire where users were asked to rate their experiences with both navigation techniques.

The four scenarios that we chose required users to navigate to a specific frame where: (1) a ball crosses an imaginary line, (2) two balls collide, (3) a ball collides with the table cushion, (4) all balls have just stopped moving.

Task completion time was measured from the first mouse-down event to when the user released the mouse button within one frame of the predetermined target frame.

Results

The data points for each video were analyzed using a paired Student's t-test (see Table 1). Users performed, on average, between 19% and 42% faster with DRAGON than with the timeline slider. In all cases, this difference was significant ($p < 0.01$).

video	length [s]	mean times [s]		% diff.	p
		slider	DRAGON		
1	3.84	3.88	2.71	32%	0.0002
2	3.52	3.67	2.86	22%	0.0077
3	7.12	3.02	2.44	19%	0.0006
4	7.08	17.80	10.18	42%	0.0052

Table 1. Mean task completion times for the videos used in our user study. The DRAGON technique performed significantly faster in all cases.

The responses collected from the questionnaire support these results – a majority of our participants preferred the object dragging technique over using the timeline slider to complete the navigation tasks. They also felt that DRAGON was quicker and easier to use (see Fig. 4).

We also collected qualitative feedback regarding the use of DRAGON for video navigation. Over two-thirds of our users agreed or strongly agreed with the statement: "DRAGON always behaved like I expected," on a five-point Likert scale. In the specific cases where the system did not behave as users expected, we were able to determine that performance issues with our prototype implementation of DRAGON played

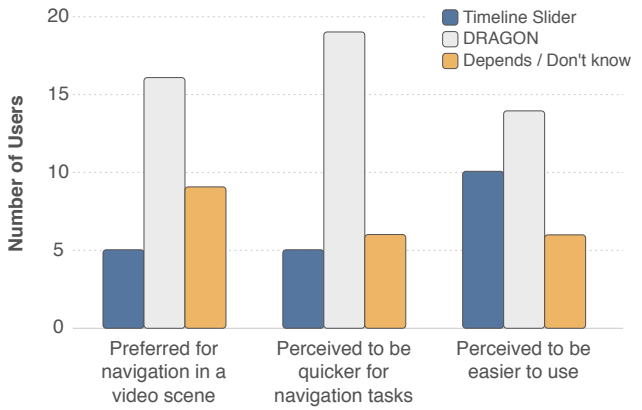

Figure 4. Results of the follow-up questionnaire on users' subjective experience with DRAGON and the timeline slider. DRAGON was preferred by the majority of the users, and was also considered both quicker and easier to use.

a major role in the usability breakdown. Two participants stated that they performed video editing tasks regularly (at least once a week), although they were not professional video editors. Both were very enthusiastic about DRAGON, and keen to see the technique incorporated in video editing software. Several other participants likewise commented that using DRAGON felt very natural. A few were even surprised when told that such interaction is not possible with current well-known video editing, browsing, or annotation software, thus indicating that the impact of a possible Hawthorne effect during the experiment was limited.

However, a majority of the participants also felt that DRAGON would not be useful for simply watching videos. Many stated that they seldom use the controls of the video player when, for example, watching a movie, and others expressed concerns that the technique was not applicable to standard home entertainment appliances such as VCRs or DVD players due to the lack of graphical input devices.

Nevertheless, about half of the participants expressed interest in having both DRAGON and the timeline slider in video player software for navigating video at varying granularities (for example, the timeline slider for navigating between scenes, and DRAGON for navigating within a particular scene). These comments coincide with our assumption that DRAGON is particularly useful for in-scene navigation, while typical navigation tasks when watching movies are more coarse-grained (such as find a particular scene). For these coarse-grained navigation tasks, adequate controls such as chapter markers and overviews already exist.

CONCLUSIONS AND FUTURE WORK

The results of our user study show how DRAGON, our direct manipulation technique for in-scene video navigation, results in a 19–42% reduction in task completion time when compared to the traditional timeline slider. Qualitative feedback from users was similarly positive, with the majority of the users preferring DRAGON over the timeline slider for in-scene navigation tasks; users also felt these tasks could be completed both more quickly and easily with DRAGON, and commented positively on its naturalness.

We are now extending DRAGON to better cope with panning and zooming in video scenes, by improving our implementation of the optical flow algorithm. We also aim to increase performance of the flow field computations so that object trajectories can be computed at runtime, thus eliminating the offline preprocessing phase. At the same time, we are experimenting with different types of visual feedback to indicate the directions in which an object can be dragged, and with modifications to the distance measure for finding the optimal frame for a given object trajectory and dragging location. Finally, we are planning a more extensive evaluation of DRAGON, comparing it to other existing techniques for improved video navigation. We would also like to evaluate how DRAGON performs in a professional video editing environment with expert users.

REFERENCES

1. Brox, T., Bruhn, A., Papenberg, N., and Weickert, J. High accuracy optical flow estimation based on a theory for warping. In T. Pajdla and J. Matas, eds., *European Conference on Computer Vision (ECCV)*. Springer, Prague, Czech Republic, 2004, volume 3024 of *LNCS*, 25–36.
2. Hürst, W. Interactive audio-visual video browsing. In *Proceedings of the MM 2006 Conference on Multimedia*. ACM Press, New York, NY, USA, 2006, 675–678.
3. Hürst, W., Götz, G., and Jarvers, P. Advanced user interfaces for dynamic video browsing. In *Proceedings of the MM 2004 Conference on Multimedia*. ACM Press, New York, NY, USA, 2004, 742–743.
4. Kimber, D., Dunnigan, T., Girgensohn, A., Shipman, F., Turner, T., and Yang, T. Trailblazing: Video playback control by direct object manipulation. *Multimedia and Expo, 2007 IEEE International Conference*, 1015–1018.
5. Lee, E. Towards a quantitative analysis of audio scrolling interfaces. In *Extended Abstracts of the CHI 2007 Conference on Human Factors in Computing Systems*. ACM Press, New York, NY, USA, 2007, 2213–2218.
6. Ramos, G. and Balakrishnan, R. Fluid interaction techniques for the control and annotation of digital video. In *Proceedings of the UIST 2003 Symposium on User Interface Software and Technology*. ACM Press, New York, NY, USA, 2003, 105–114.
7. Shneiderman, B. Direct manipulation: A step beyond programming languages. *IEEE Computer*, *16*, 8 (1983), 57–69.
8. Shneiderman, B. and Plaisant, C. *Designing the User Interface*. Addison Wesley, 2005, 4th edition.
9. Zhai, S., Smith, B. A., and Selker, T. Improving browsing performance: A study of four input devices for scrolling and pointing tasks. In *Proceedings of INTERACT 1997 Conference on Human-Computer Interaction*. Sydney, Australia, 1997, 286–292.