Improving Video Browsing with an Eye-Tracking Evaluation of Feature-Based Color Bars

Neema Moraveji
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA
+1-412-268-4050
neema@cmu.edu

Abstract

This paper explains a method for leveraging the standard video timeline widget as an interactive visualization of image features. An eye-tracking experiment is described with results that indicate that such a widget increases task efficiency without increasing complexity while being easily learned by experiment participants.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]

General Terms

Design, Experimentation, Human Factors.

Keywords

Digital video, user interface.

1. INTRODUCTION

Digital video is fast becoming a common medium for dissemination via download and home movies. For example, end-users commonly download music videos from web sites and watch documentaries housed in digital libraries. Automatic feature classifiers are ameliorating the dearth of metadata and user interface enhancements leveraging this metadata are being created to improve video information foraging. However, the problem of video information retrieval remains unsolved. Part of the problem is finding the most relevant video segment from a large set. Another problem, the one this paper focuses on, is efficiently browsing a video segment that contains a relevant sequence.

An interactive timeline is standard on most, if not all, media players. While query-based video retrieval interfaces can show users where matches occur in the video document (e.g. red vertical stripes on the timeline), the image feature context surrounding that match is lost. Furthermore, when few or no matches are in a video, browsing a video with the timeline becomes an exercise in trial-and-error. By leveraging image features, these shortcomings may be addressed while increasing task efficiency. This paper describes the design and eye-tracking evaluation of a timeline enhanced with image feature data.
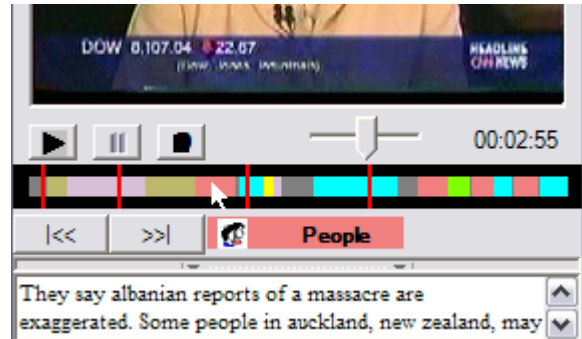
2. FEATURE-BASED COLOR BARS

Unique and visually distinct colors are assigned to image features and the timeline is painted according to the most relevant feature found in that portion of the segment, effectively transforming the timeline into a set of "color bars" [1] (Figure 1). Colors are "visually distinct, can codify a large number of categories, and are identifiable even in small slices" [1]. However, bars that are too narrow to easily click are removed from the set. When the user places the cursor over a bar, a graphical icon and text descriptor of that feature is displayed immediately below the timeline. This allows users to learn new features quickly and compensates for the fact that users will likely forget what each colors means. It should be noted that any scene in a video document can have multiple features associated with it. In this design, one feature per scene was chosen to reduce clutter, delineating scenes intuitively. Clicking on a colored area of the timeline moves the video to that portion of the video segment.

3. EXPERIMENT

Twelve subjects (4 male) with an average age of 23 were each given 4 tasks with order randomized. Users listened to an audio tutorial of the interface controls and spent several minutes having the head-mounted eye-tracker calibrated to their visual angle. For each task, the description was shown along with a media player that consisted of the video image, playback controls, a timeline, buttons to navigate between query matches, and a text transcript that allowed users to browse the segment by clicking on the corresponding text. Each task displayed to the user one of the following treatments:

- timeline with colors bars and query match lines ('Both')
- timeline with only color bars ('Color')
- timeline with only query match lines ('Lines')

- timeline with no extraneous information ('Neither')

Subjects navigated a video document until they found a correct answer and submitted their answer by voice to the conductor. Tasks were created in accordance with the video browsing assessment metrics proposed by Wildemuth, et al [5]. Textual tasks described the task in words while the graphical task presented a digital image taken from the video and asked users to find the scene that the image was taken from. Tasks also had a mixture of unique and multiple answers:

- "Find a person ABNews interviewed in Iraq."
  (A - *Action recognition* – textual, multiple possible answers)
- "Find a scene depicting text written on trophy basketballs."
  (B - *Object recognition* – textual, one answer)
- "Find the scene where the head of Microsoft sings a song."
  (C - *Action recognition* – textual, one answer)
- "Find an image of this clean-burning car."
  (D - *Object recognition* – graphical, one answer)

The TRECVID 2003 FSD corpus [4] was used to pre-select the video returned to the user for each task, and the actual query words used to find that video were displayed under the task description. The mean video segment duration was 5:39, ranging from 4:19 to 7:24. This duration was chosen to make watching the entire segment an inefficient method of completing the task. Relevant image features were chosen from TRECVID 2002 and 2003 and manually coded to be relevant to the task at hand. For example, for tasks whose goal it was to find a person, the Face feature took precedence over the Vehicle feature.

4. RESULTS

Color bars improved task efficiency on all tasks, but at different rates, ranging from 18.4% (task C) to 234.1% (task D). The uniqueness of the feature most relevant to the task had a large impact on how helpful the color bars were. However, even for tasks where the most relevant feature was common and distributed evenly throughout the segment, color bars improved efficiency noticeably.

The number of times the two browsing widgets (timeline and transcript) were used was also recorded. Without color bars, the timeline was used 32% of the time to browse the video. When color bars were available, timeline use soared to 69%.

4.1 Eye-tracking Data Analysis

We inspected the relative amount of fixations (200 ms threshold [1]) that the timeline attracted during each treatment, assuming that this metric "reflects the importance of that element" [2]. The number of fixations overall, believed to be negatively correlated with search efficiency [2], was also measured for each

Table 1: Eye-tracking metrics for the four treatments.

	Mean % attention on timeline	Mean % fixations on timeline	Total fixations
Neither	10.7%	12%	112.1
Lines	17.7%	17%	114.5
Color	35.2%	36%	78.8
Both	34.9%	28%	64.83
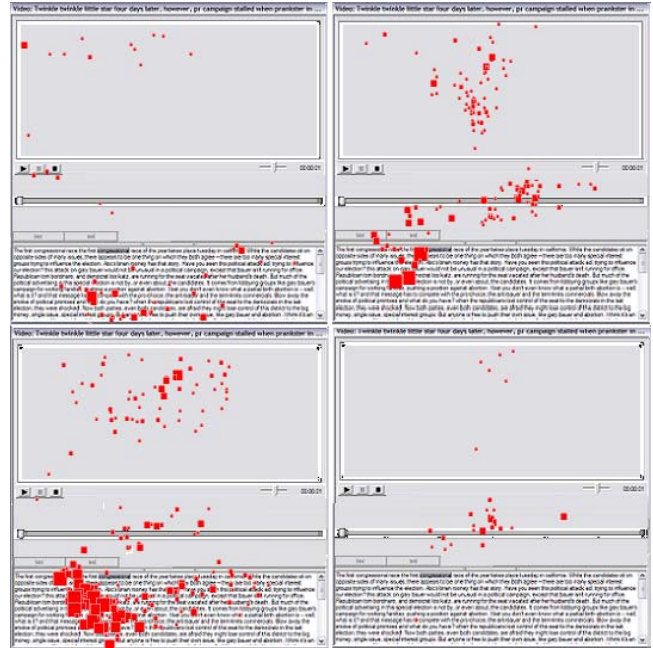
Figure 2: Representative subject's fixations across four treatments shows emphasis on timeline when color is visible. Clockwise from top left: Neither, Color, Both, Lines.

topic to ensure that adding more visual information didn't increase the complexity of the interface. Results indicate that color bars increased the portion of the user's attention paid to the timeline while decreasing the average number of fixations required to solve the task (Table 1). A visual comparison of the fixations that a representative subject made in all four treatments is presented in Figure 2. The size of a plot on the figure correlates to the duration of that fixation.

5. CONCLUSION AND FUTURE WORK

Placing feature-based color bars on a media player timeline is shown to be an efficient interaction technique enabling informed browsing of a video segment. Subject quickly learned how to use the widget and found it to be more useful for browsing than both the segment transcript and an undecorated timeline. Future work will investigate giving users control over what features are displayed, considering the ramifications of multiple features displayed in the same temporal range, and examining the effectiveness of the widget with imperfect feature classifiers.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] A. Hughes, et al (2003). "Text or Pictures? An Eye Tracking Study of How People View Digital Video Surrogates." CIVR.

[2] R. Jacob, K. Karn (2003). "Eye tracking in human-computer interaction and usability research." Oxford: Elsevier Science.

[3] H. Liu, T. Selker, H. Lieberman (2003). "Visualizing the affective structure of a text document." SIGCHI.

[4] TRECVID 2003 Guidelines. http://www-nlpir.nist.gov/projects/tv2003/tv2003.

[5] B. M. Wildemuth, et al. (2003). "How Fast Is Too Fast? Evaluating Fast Forward Surrogates for Digital Video" JCDL.