

SILVER: Simplifying Video Editing With Metadata

A. Chris Long¹, Juan Casares, Brad A. Myers, Rishi Bhatnagar, Scott M. Stevens,
Laura Dabbish, Dan Yocum, and Albert Corbett

Human Computer Interaction Institute

Carnegie Mellon University

Pittsburgh, PA 15213 USA

+1 412 268 3527

chrislong@acm.org

ABSTRACT

Digital video is becoming increasingly ubiquitous. However, editing video remains difficult for several reasons: it is a time-based medium, it has dual tracks of audio and video, and current tools force users to work at the smallest level of detail. Based on interviews with professional video editors, we developed a video editor, called Silver, that uses metadata to make digital video editing more accessible to novices. To help users visualize video, Silver provides multiple views with different semantic content and at different levels of abstraction, including storyboard, editable transcript, and timeline views. Silver offers smart editing operations that help users resolve the inconsistencies that arise because of the different boundaries in audio and video.

Keywords

Digital video editing, multimedia authoring, metadata, Silver, Informedia

INTRODUCTION

Digital video is becoming increasingly ubiquitous. As technology and content become available, users will shift from passive spectators to active creators of video. Forrester Research predicts that by 2005, 92% of online consumers will create personal multimedia content at least once a month [4]. However, although many exciting research projects are investigating how to search, visualize, and summarize digital video, there is little work on new ways to support the use of the video beyond just playing it.

This is unfortunate, because video editing has several unique challenges not found with other media. One is that digital video is a time-based medium, which hampers browsing and skimming video.

Another challenge for editing video is that it is a dual medium, consisting of an audio track in addition to the video track. These tracks must be kept synchronized, but the user must also be able to overlay them when desired, for example during transitions from one shot to another. Further, when a shot is cut for use elsewhere, the user must be able to disentangle overlaid audio and video.

A third problem is that the syntactic units that users want to edit are shots of video and words or sentences of audio, but current tools require users to examine video at the individual frame level and audio using a waveform. To perform most editing operations, the user must manually pinpoint specific frames, which may involve zooming and many repetitions of fast-forward and rewind operations.

These problems make editing video a difficult, tedious, and error-prone activity. Commercially available tools, such as Adobe Premiere [1] and Apple iMovie [2], allow creation of high-quality video, but they do not adequately address the issues raised above, which makes them harder to use, especially for novices.

Based on observations of an interviews with video professionals, we created Silver, a system that uses metadata and video-specific interaction techniques to make video editing easier [3]. The key innovations in the Silver editor include: providing an editable transcript view; coordinating the selection across all views including partial selections and different selections in audio and video; and smart editing through smart snap, smart selection and smart cut and paste.

Silver simplifies video editing primarily by using metadata such as shot breaks, thumbnail images, and clip titles—all of which are obtained from the Informedia digital library [5]—to enhance the display of and interaction with the video composition, and by providing multiple, synchronized views of the video composition.

SILVER INTERFACE

An overview of the Silver interface, with all the views it offers, is shown in Figure 1. At the highest level of

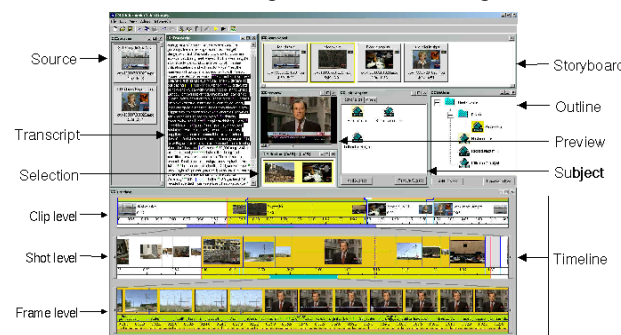


Figure 1 Overview of the Silver interface

Copyright is held by the author/owner(s).

CHI 2003, April 5-10, 2003, Ft. Lauderdale, Florida, USA.

ACM 1-58113-637-4/03/0004

¹Current affiliation: ECE Department, Carnegie Mellon University.

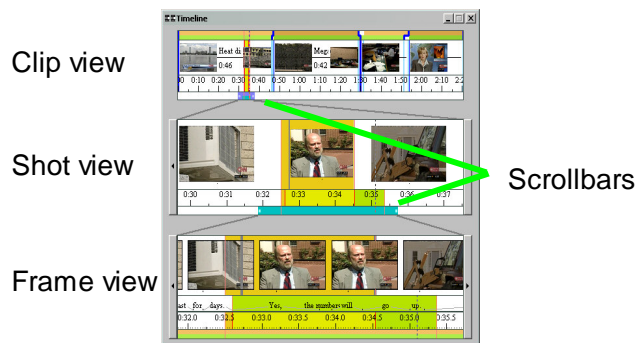


Figure 5 Timeline window.

abstraction, the composition is composed of video *clips* (i.e., video from a single file). Clips that have been previously found are shown in the source window. Clips can be organized in the outline and subject views, and can be put in the composition by dragging into in the storyboard window, where they can also be rearranged. When part of the composition is selected, the selection window shows the starting and ending frames.

The main part of the Silver interface is the hierarchical timeline, which shows video at three scales: clip, shot, and frame (see Figure 2). The clip and shot levels show the video composition as continuous segments from a single source file or camera, respectively. Each clip is displayed as a representative frame and, if space allows, its end frame, title, and duration. Each shot is displayed using a representative frame.

Users control how much is displayed in the shot and frame levels using a sophisticated scrollbar at the bottom of the clip level. In addition to the usual dragging behavior, the scrollbar may also be resized to zoom the shot level in or out. It also contains a miniature version of the shot level scrollbar, which shows the position of the frame level within the shot level. As in the clip level, there is an augmented scrollbar at the bottom of the shot level used to scroll and zoom the frame level.

The frame level displays the individual frames of the video, so the user can quickly find particular cut points. The frame level addresses the dual medium issue by displaying the transcript, including the start and end times of every word.

An important feature of the timeline is smart selection and editing. Silver uses metadata to ease editing with semantically meaningful units, for example, by snapping selection to shot boundaries and audio selection to word boundaries (see Figure 2). The user can disable snapping for manual selection if desired.

Another key feature of the timeline is that it shows different representations together, allowing the user to see the clip boundaries, the time, samples of frames, the cuts, the transcript, annotations, etc.

Silver also displays a separate, synchronized view of the transcript in another window (see Figure 3). Selection in

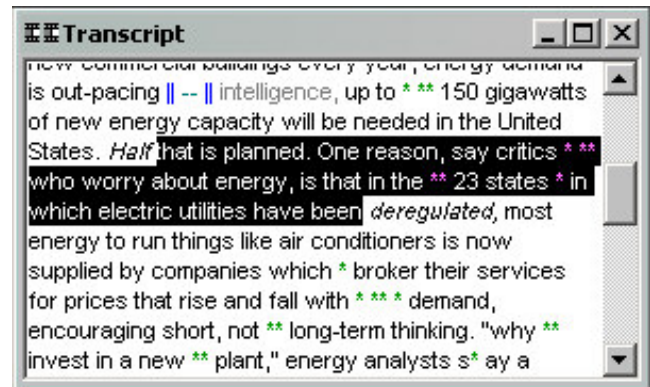


Figure 4 Transcript window. Blue vertical bars ("||") show clip breaks. Dashes show gaps in audio. The word "intelligence" was cut in the middle. The green stars represent noise or recognition errors. At the edges of the selection, *half* and *deregulated* are partially selected.

the transcript and timeline are synchronized, and editing the transcript edits the video composition.

Silver allows audio and video to be selected independently. However, inserting or deleting unequal amounts of audio and video creates a hole in one of the tracks. When this happens, Silver notifies the user and provides options for closing the gap.

CONCLUSIONS

Silver is designed to make high-quality video editing easier, and demonstrates several techniques for overcoming difficulties of authoring in this domain. Specifically, it uses coordinated multi-level views, synchronized selection, and a textual transcript.

Silver shows that video metadata can be used to create an advanced interface for aiding novice users in the performance of otherwise difficult or tedious tasks. As more metadata becomes available, the smart editing approach of Silver can be used to improve video editing capabilities even further.

ACKNOWLEDGMENTS

The Silver Project is funded in part by the National Science Foundation under Grant No. IIS-9817527, as part of the Digital Library Initiative-2. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the National Science Foundation.

REFERENCES

1. Apple Computer, Inc. iMovie. <http://www.apple.com/imovie/>.
2. Adobe Systems Incorporated. Adobe Premiere. <http://www.adobe.com/products/premiere/main.html>.
3. Casares, J., Long, A. C., Myers, B. A., Bhatnagar, R., Stevens, S. M., Dabbish, L., Yocum, D., and Corbett, A. "Simplifying Video Editing Using Metadata." In *Proceedings of DIS2002* (London, UK, June 2002), p. 159–166.
4. Schwartz, J., Rhinelander, T. and Dorsey, M., "Personal Rich Media Takes Off", The Forrester Report, Forrester Research Inc., October 2000.
5. Wactlar, H.D., et al., "Lessons learned from building a terabyte digital video library.", *IEEE Computer*, 1999. 32(2): pp. 66-73.