# Video Metadata Viewer

**Mackenzie Leake**
Stanford University
Stanford, CA USA
mleake@cs.stanford.edu

## ABSTRACT
Abstract text . . .

### ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous; See http://acm.org/about/class/1998/ for the full list of ACM classifiers. This section is required.

### Author Keywords
Authors' choice; of terms; separated; by semicolons; include commas, within terms only; required.

## INTRODUCTION
Introduction text . . .

## RELATED WORK

### Browsing and summarizing videos
To support browsing content within videos many commercial video editing and playback interfaces require users to scrub through the timeline to find relevant sections. Goldman et al., 2006 present semantic storyboards as an alternative technique to timeline scrubbing for browsing videos [7]. Using a static image and the visual language of storyboards, the system provides users with an overview of the video content. Barnes et al., 2010 [2] present another alternative to video scrubbing through the creation of video tapestries, which provide multiscale image summaries of videos. Zhang, Wang, and Altunbasak, 1995 [13] use 2-D key-frame images to support users in browsing videos. Instead of using static 2-D images to represent video content, Panopticon, a video alternative system, presents an overview of the full video content to users by displaying an animated grid of video subsequences [8]. Like these systems my project reduces the need for scrubbing through videos to find relevant content. My system will provide additional mechanisms for exploring a video's underlying metadata.

Truong and Venkatesh, 2007 [11] discuss the importance of abstraction for video browsing and summarization. Silver, a system designed by Casares et al., 2002, provides multiple views, including storyboard, transcript, and timeline views, to support the editing process [4]. Video digests, presented by Pavel et al., 2014, support the browsing and skimming of instructional videos by pairing video segments with their summaries [10]. The video digest system allows users to create browsable video digests using transcripts, chapter summaries, and keyframes. While Silver and the video digest system consider video at various levels of abstraction, including the chapter, clip, shot, and frame levels, my project primarily visualizes clips at a different level of abstraction: the line level. This decision emphasizes a closer connection between the transcript content and the video clip.

SceneSkim is a system that supports the alignment of the script, plot summary, and captions to videos for browsing content in feature films [9]. This system utilizes the alignment between videos and text elements for studying existing films. My project applies many of these techniques for editing new video sequences.

### Video editing
The IMPACT system proposed by Ueda, Miyatake, and Yoshizawa, 1991 [12] addresses the difficulty of editing video by presenting the use of image processing techniques to support multimedia editing. Hitchcock, the automatic home editing system proposed by Girgensohn et al., 2000, utilizes camera motion and video editing conventions regarding clip length and brightness to identify suitable clips for inclusion in the final edited sequence [6]. My project also makes use of video metadata for supporting the automatic editing process. It makes an additional connection between videos and the spoken narrative by presenting the user with video clips aligned to their corresponding transcripts.

DemoCut is a system for semi-automatically editing instructional videos for physical demonstrations [5]. Users identify important steps in the video, and then the system automatically sequences the video and audio segments. Arev et al., 2014 present methods for automatically editing videos taken from multiple social camera based upon cinematographic principles and important actions in the scene [1]. My project, too, encourages users to be mindful of incorporating a variety of shot types into their edited sequences, but it also uses the alignment of a video with its transcript to guide the editing process.

Berthouzoz, Li, & Agrawala, 2012 present a set of tools for placing smooth transitions in interview videos [3]. These tools make use of the alignment between the video and its transcript

in order to guide automatic edits. My project is similar in its emphasis on interview-style video and use of transcript-video alignment. My project, however, also presents the user with additional information, such as identifying the emotional content in the script, to support the user in understanding the available video content for editing.

## PROJECT PLAN

In Fig. 1 I show a sketch of the interface I am planning to build:

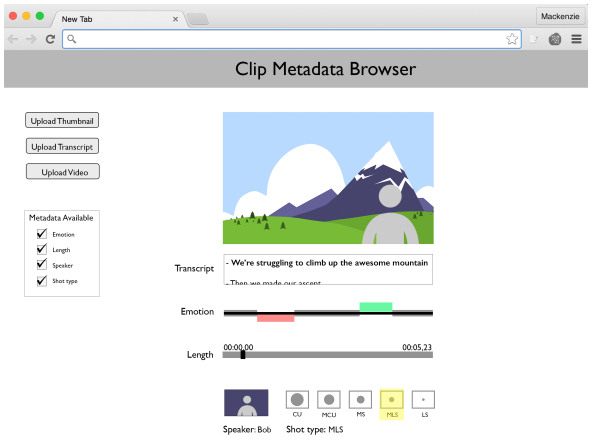


**Figure 1. Current interface design.**

Currently I have the following features implemented:

- A basic web page with header
- Three menu buttons for uploading a thumbnail image, a video, and a transcript
- Two sets of videos, transcripts, and additional metadata, i.e., spoken content, emotion scores, line length, speaker, and shot type.

Next steps include:

- Data handling (by May 26)
  - Condense and format metadata into a single JSON for input.
  - Parse the JSON.
  - Pass appropriate data to separate functions to build sub-visualizations.
- Design layout of webpage (by May 30).
- Design sub-visualizations (by June 1).
  - Scrolling transcript browser
  - Emotion score viewer
  - Line length bar
  - Speaker thumbnail box
  - Shot type selection

## SYSTEM OVERVIEW

System overview text . . .

## RESULTS

Results text . . .

## DISCUSSION

Discussion text . . .

## CONCLUSION

Conclusion text . . .

## ACKNOWLEDGMENTS

Acknowledgments text . . .

## REFERENCES

1. Ido Arev, Hyun Soo Park, Yaser Sheikh, Jessica Hodgins, and Ariel Shamir. 2014. Automatic editing of footage from multiple social cameras. *ACM Transactions on Graphics (TOG)* 33, 4 (2014), 81.

2. Connelly Barnes, Dan B Goldman, Eli Shechtman, and Adam Finkelstein. 2010. Video tapestries with continuous temporal zoom. In *ACM Transactions on Graphics (TOG)*, Vol. 29. ACM, 89.

3. Floraine Berthouzoz, Wilmot Li, and Maneesh Agrawala. 2012. Tools for Placing Cuts and Transitions in Interview Video. *ACM Trans. Graph.* 31, 4, Article 67 (July 2012), 8 pages. DOI: http://dx.doi.org/10.1145/2185520.2185563

4. Juan Casares, A. Chris Long, Brad A. Myers, Rishi Bhatnagar, Scott M. Stevens, Laura Dabbish, Dan Yocum, and Albert Corbett. 2002. Simplifying Video Editing Using Metadata. In *Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques (DIS '02)*. ACM, New York, NY, USA, 157–166. DOI: http://dx.doi.org/10.1145/778712.778737

5. Pei-Yu Chi, Joyce Liu, Jason Linder, Mira Dontcheva, Wilmot Li, and Bjoern Hartmann. 2013. DemoCut: Generating Concise Instructional Videos for Physical Demonstrations. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. ACM, New York, NY, USA, 141–150. DOI: http://dx.doi.org/10.1145/2501988.2502052

6. Andreas Girgensohn, John Boreczky, Patrick Chiu, John Doherty, Jonathan Foote, Gene Golovchinsky, Shingo Uchihashi, and Lynn Wilcox. 2000. A semi-automatic approach to home video editing. In *Proceedings of the 13th annual ACM symposium on User interface software and technology*. ACM, 81–89.

7. Dan B Goldman, Brian Curless, David Salesin, and Steven M Seitz. 2006. Schematic storyboarding for video visualization and editing. In *ACM Transactions on Graphics (TOG)*, Vol. 25. ACM, 862–871.

8. Dan Jackson, James Nicholson, Gerrit Stoeckigt, Rebecca Wrobel, Anja Thieme, and Patrick Olivier. 2013. Panopticon: A Parallel Video Overview System. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. ACM, New York, NY, USA, 123–130. DOI: http://dx.doi.org/10.1145/2501988.2502038

9. Amy Pavel, Dan B. Goldman, Björn Hartmann, and Maneesh Agrawala. 2015. SceneSkim: Searching and Browsing Movies Using Synchronized Captions, Scripts and Plot Summaries. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM, New York, NY, USA, 181–190. DOI: http://dx.doi.org/10.1145/2807442.2807502

10. Amy Pavel, Colorado Reed, Björn Hartmann, and Maneesh Agrawala. 2014. Video Digests: A Browsable, Skimmable Format for Informational Lecture Videos. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*. ACM, New York, NY, USA, 573–582. DOI: http://dx.doi.org/10.1145/2642918.2647400

11. Ba Tu Truong and Svetha Venkatesh. 2007. Video Abstraction: A Systematic Review and Classification. *ACM Trans. Multimedia Comput. Commun. Appl.* 3, 1, Article 3 (Feb. 2007). DOI: http://dx.doi.org/10.1145/1198302.1198305

12. Hirotada Ueda, Takafumi Miyatake, and Satoshi Yoshizawa. 1991. IMPACT: an interactive natural-motion-picture dedicated multimedia authoring system. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 343–350.

13. HongJiang Zhang, Chien Yong Low, Stephen W Smoliar, and Jian Hua Wu. 1995. Video parsing, retrieval and browsing: an integrated and content-based solution. In *Proceedings of the third ACM international conference on Multimedia*. ACM, 15–24.