

Video Metadata Viewer

Mackenzie Leake
Stanford University

How can we help people produce better edited videos?

Motivation

Editing videos is challenging because:

- The amount of video content is often overwhelming.
- Video has many levels of abstraction (Fig. 1).
- Visual and audio elements must be combined into a coherent narrative.

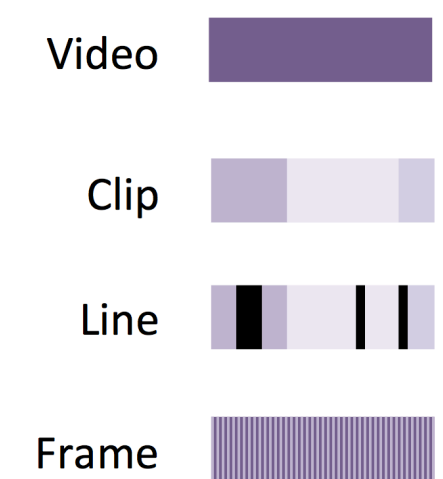
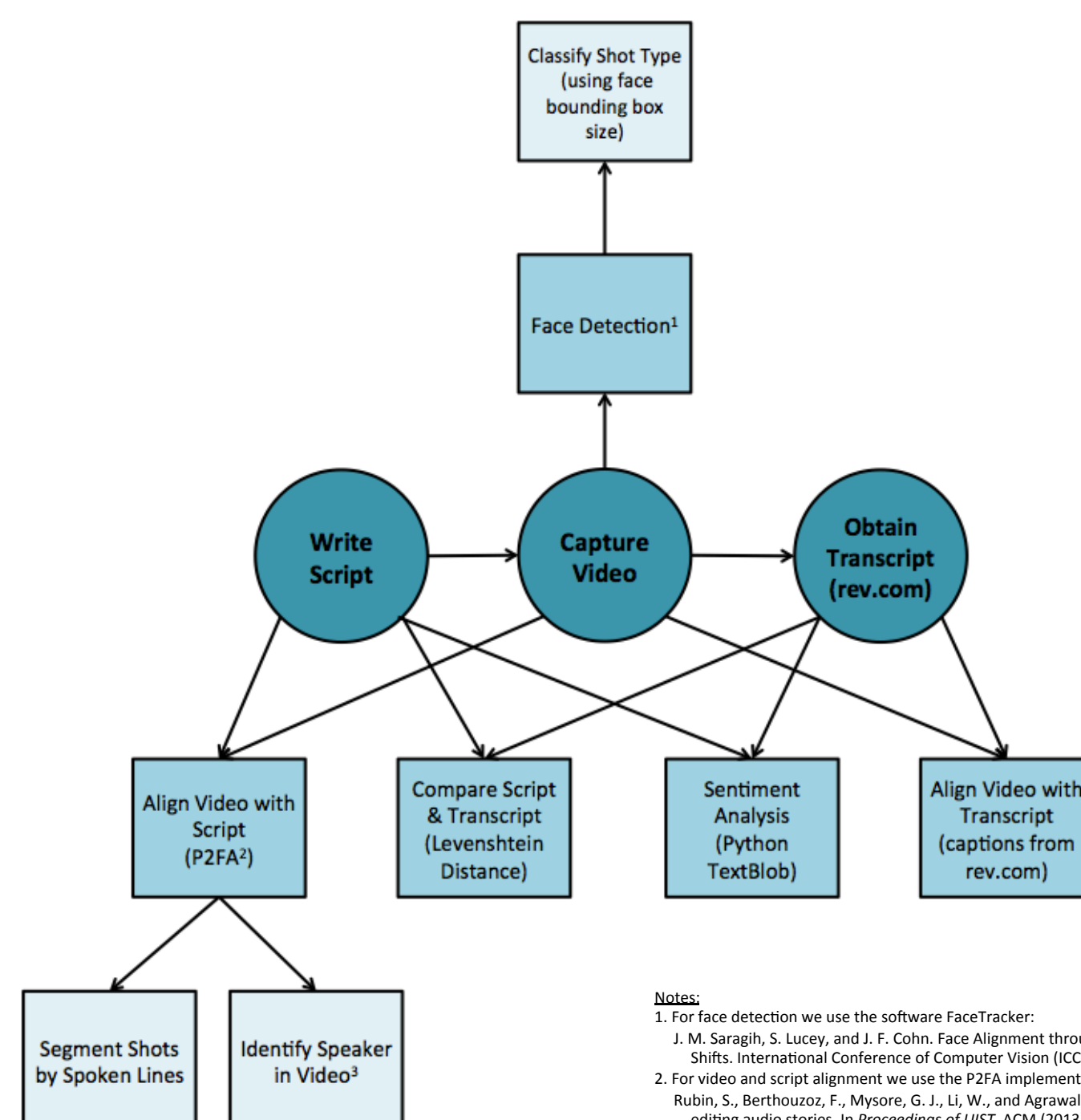


Figure 1. Levels of video abstraction.

Approach

We film a scripted scene from multiple camera angles and obtain data from these videos using the following process:



Notes:
1. For face detection we use the software FaceTracker:
J. M. Saragih, S. Lucey, and J. F. Cohn. Face Alignment through Subspace Constrained Mean-Shifts. International Conference of Computer Vision (ICCV), September, 2009.
2. For video and script alignment we use the P2FA implementation discussed in:
Rubin, S., Berthouzo, F., Mysore, G. J., Li, W., and Agrawala, M. Content-based tools for editing audio stories. In Proceedings of UST, ACM (2013), 113–122.
3. For speaker detection we currently use the script label, but in future work we plan to use face detection for this labeling.

Results: Line Representation

Line Level Features:

- 1 Video player: watch video
- 2 Shot selector: check video filenames
- 3 Script & transcript viewers:

A. Visual encodings:

- I. Rectangle width represents line length.
- II. Hue indicates speaker.
 - Anna
 - Doctor
- III. Opacity indicates the emotion of the line.

B. Interactions:

- I. Hover over a rectangle for line text
- II. Click on a rectangle to go to a line in the video

Results: Frame Representation

Frame Level Features:

- 1 Frame viewer: display face detection points
- 2 File selector: select video to examine
- 3 Frame selector: slide to select frame
- 4 Speaker information: list actor speaking
- 5 Shot type indicator: provide shot zoom type

Future Work

- Develop upload interface to allow user input.
- Evaluate trade-offs of line versus frame level representations of videos.
- Design editing interface.