AutoTag: Metadata Automation for Film Post-Production

José Marcelo Sandoval-Castañeda

Co-advisors: Dennis Shasha and Scandar Copti
Presentation supervisors: Dennis Shasha and Nizar Habash

May 12th, 2020

Film Production Process

General Overview

- Pre-production: planning, writing, hiring.
- Production: the actual shooting of the film.
- **Post-production**: editing.
  - **Metadata tagging**.
  - Actual video editing.
  - ADR/Foley.
  - Special effects.

Metadata Tagging

And why it is extremely inefficient

- Shooting ratios tend to be between 100:1 and 1000:1.
- Usually done by hand.
- **Identify shot types**.
- Identify relevant objects.
- **Identify scenes**.
- **Transcribe dialogue/interviews**.
- Determine quality.

AutoTag's Functionality

Main Goals

- Provide a workable and searchable **transcript** of media files.
- Identify the **scenes** that contain the media files transcribed.
- Identify the **type of shot** of the video files.
- Output all of this information in Adobe Premiere Pro.

Transcribing Dialogue

For both fiction and documentary films

- Google Cloud’s Speech-to-Text services.
- Audio from media files is transcribed 10 seconds at a time.
- Data is output as **markers** within Adobe Premiere Pro:
  - Searchable.
  - Readable.
  - Can be exported into other NLE software if needed.

Identifying the Scene

For fiction films

- Extraction of dialogue from a screenplay file.
- Stemming and construction of **n-grams** (2 to 5) from the transcript of the media files and the screenplay's scenes.
- **Jaccard similarity vectors** for each media file to all scenes.
- Highest Jaccard similarity is the scene assigned to the media file.
- Scene is assigned as **metadata** information in Adobe Premiere Pro.
- **Search bins** for each scene are created.

Identifying the Shot Type

For both fiction and documentary films

- Face detection performed by a ResNet-based **Single Shot Detector**.
- **K-Means clustering** with five centroids for more than 1 million frames extracted from 12 different movies.
- Each cluster is assigned a name: close-up, medium close-up, medium, American, and long.
- Shot type is assigned as **metadata** information in Adobe Premiere Pro.
- **Search bins** for each shot type are created.

(a) A close-up shot. Usually includes the subject's head and neck in a tight frame.

(b) A medium close-up shot. Usually goes from the subject's shoulders to the top of the head.

(c) A medium shot. Usually goes from the subject's torso or hips to the top of the head.

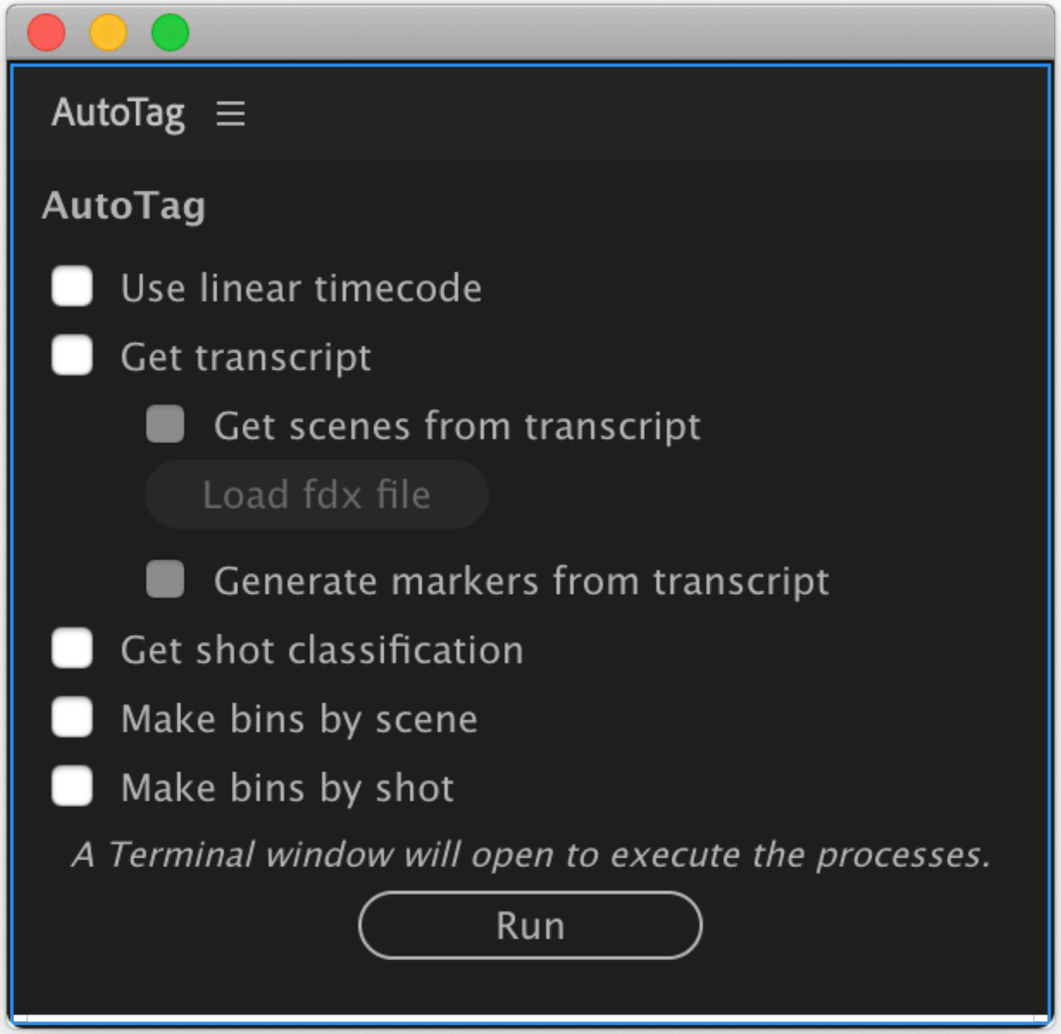
(d) An American shot. Usually goes from the subject's knees to the top of the head.

(e) A long (or wide) shot. Usually includes the full body of the subject, and arbitrary amounts of head space.

Interface

Inside Adobe Premiere Pro

- Toggle-able options for each step in the process.
- Transcript-dependent options are indented and disabled unless the user also extracts the transcript.
- Run button launches a Terminal window where a progress bar is displayed.



Results

From two case studies

Case Studies	*Challenging Erasure* by Katarina Holtzapple	*How We Leave* by Liene Magdalēna
Shooting Ratio	80.7:1	58.6:1
Median Time for Transcription	3.21 seconds for every 10 seconds	4.05 seconds for every 10 seconds
Scene Identification	Not applicable	90.3% accuracy
Shot Type Classification	91.8% accuracy	88.2% accuracy

Testimonials

From the filmmakers

- Katarina Holtzapple:
  “This tool is particularly useful for the work I do […] because so much of my work is found footage and **shots taken for over a year** and that can get very disorganized.”

  She estimates that at least **25% of her post-production time** was spent on tasks performed by AutoTag.

- Liene Magdalēna:
  “A tool like this allows us filmmakers to **focus on the art involved in post-production**, not having to worry much about the more technical and repetitive tasks.”

  Her post-production process took about **5 months**, a significant portion of which was wrestling with metadata and communication across editors.

Future Work

- Expand its compatibility to **other NLE software**, such as Blackmagic's DaVinci Resolve and Avid Media Composer.
- Perform **more case studies** with more filmmakers and on larger projects.
- Improve on the **algorithms** used: identification of speakers, motion identification, supervised learning approach.
- Translate the processing of files to Javascript to allow for smoother and potentially faster metadata automation.