MIDB: A Web-Based Film Annotation Tool

Azzam Althagafi
Univ. of Pennsylvania
aazzam@seas.upenn.edu

Hui-Yin Wu
North Carolina State Univ.
huiyin_wu@ncsu.edu

Arnav Jhala
North Carolina State Univ.
ahjhala@ncsu.edu

Abstract

We present the Movie Insights Database (MIDB) annotation tool that allows for a web-based upload and annotation of films for efficient film-data collection. MIDB is light-weight and can be installed on a web server for easy access. The tool provides automated scene boundary detection, an interface for defining the annotation language, upload and download of new or existing annotations, shot boundary adjustment, and tools for quick and efficient shot-by-shot annotation. Our primary motivation for the development of this tool is to create an easily accessible platform for video analysis that allows AI assisted annotation.

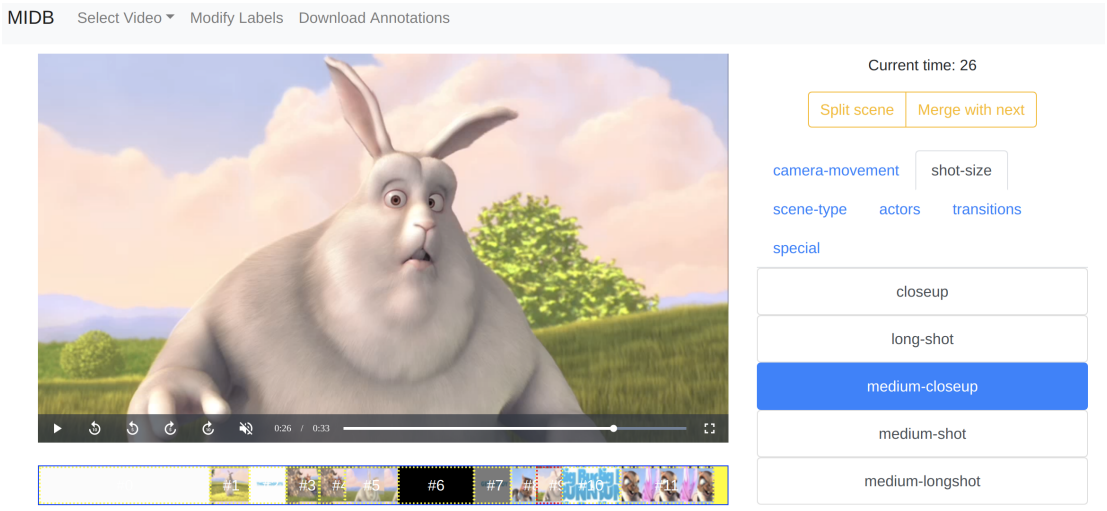

Figure 1: The MIDB annotation tool has been designed to be simple and accessible from a browser. Functions include file upload and label customization menus, a video player on the left with standard time controls, shot segmentation functions and display in the form of a timeline, and panels for different label categories available for annotation. Screenshots © copyright 2008, Blender Foundation / www.bigbuckbunny.org

In: H. Wu, M. Si, A. Jhala (eds.): Proceedings of the Joint Workshop on Intelligent Narrative Technologies and Workshop on Intelligent Cinematography and Editing, Edmonton, Canada, 11-2018, published at http://ceur-ws.org

1 Background

For almost 15 years, the Cinemetrics annotation tool and database (www.cinemetrics.lv) has collected over 19 thousand individual annotations on the cutting tempo and cinematographic features of films from all ages. The tool is particularly popular due to two reasons: (1) it is easy to use directly in parallel with a film playing on the side, allowing the user to quickly tag shots and transitions using the keyboard, and (2) it is widely accessible with both desktop and online applications that offer direct uploading and recording to its database. The database and tool is widely popular among film analysts to gain statistics for film pacing.

The recent appearance of film annotation tools such as Advene [AP05], Insight [MWS+15], and Anvil [Kip12] shows a growing trend of drawing observations on film data from a data scientist point of view. The tools provide much more advanced functionalities such as accurate timelines, automatic face detection, 3D head models, color analysis, etc. with the goal of increasing both the quality and the dimensions of the collected data. However, these tools have not been broadly accepted by the film community, and compared to Cinemetrics, have much fewer users and thus fewer annotations. This can be attributed to the fact that some of these tools require time, and possibly technical knowledge, to install and to familiarize with their various functionalities.

To this end, we propose the MIDB, which was inspired by the Insight annotation tool. It has integrated shot boundary detection using *ffmpeg*, intuitive navigation tools (by frame, by shot, and video player controls), and customizable side panels for specifying annotation tags. Moreover, for accessibility, MIDB is written completely in javascript, which makes it easy to expand its functionality with plugins, and can be set up on a web server or installed independently on a personal computer.

2 MIDB Tool

In this section we introduce the architecture, workflow, and interface design of the MIDB tool.

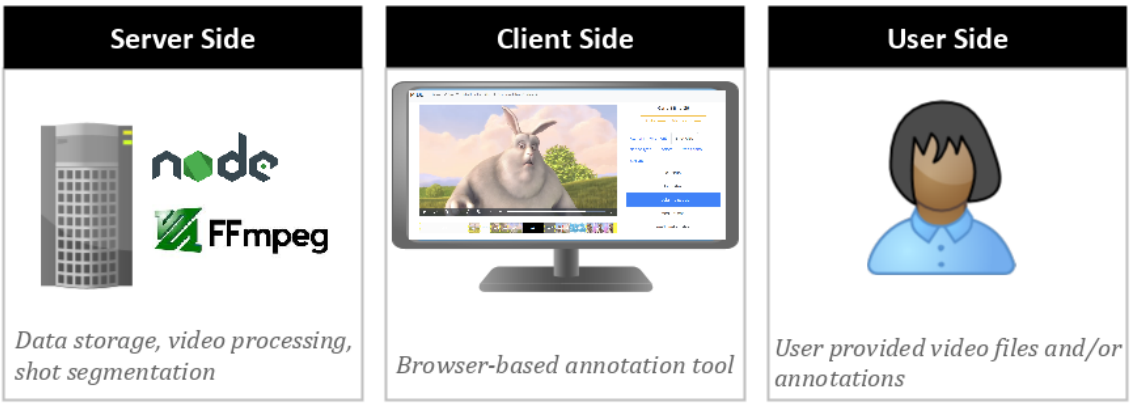

Figure 2: The framework of the MIDB annotation tool consists of the server-side storage and video processing tools, client-side browser-based interface using the javascript React framework, and user-side provided video and annotation input.

2.1 Architecture

The tool uses *node.js* for backend, and React framework for the front end interface.

Our chosen tool for shot boundary detection is *ffprobe*, which is a multimedia analysis toolkit that is integrated and makes use of the *ffmpeg* libraries.

Other plugins that are currently under development for MIDB include a trained model for automatic shot size assignment, and automatic actor detection. In the future, we also expect to experiment plugins for more advanced object and camera movement to further automated the annotation process.

The user annotations are saved and immediately backed up in *json* format that is supported by most modern programming languages.

The overview of the framework can be seen in Figure 2.

2.2 User Interface

Our user interface resembles the Insight annotation tool in its design. Using the React framework, the video player is smoothly integrated with frame-level and shot-level player controls, a timeline with shot segmentation, and a panel displaying the categories for annotation labels. A navigation bar on top allows the uploading of a video or an existing *json* annotation, and the download of annotations. Moreover, popup menus are designed to allow customization of the labels for annotation, and the interface scales and adjusts according to different screen sizes, such as by stowing the menu bar.

The tool is shown in Figure 1, showing the example annotation of the shot size (i.e. how close the camera is to the actor) in the labels panel. The side-by-side design of the video, synchronized with the display of annotation labels and timeline, make it possible to be highly precise in terms of attributing the labels to a specific shot. It also makes changing and removing annotations easy, and all user operations are immediately recorded.

2.3 Workflow

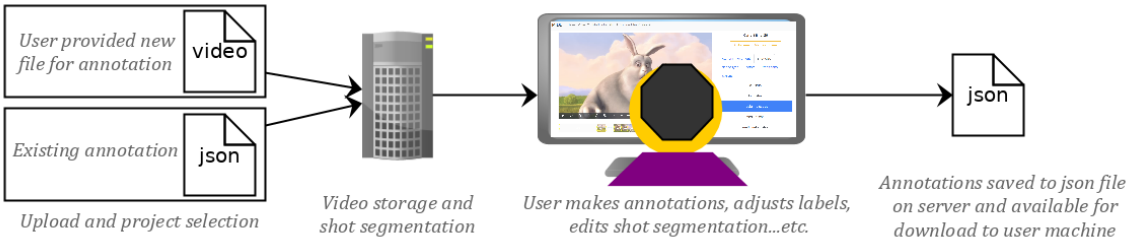

Figure 3: The workflow of the tool is simple, consisting of initiating a new or existing project, video processing (taken care of by the server, user annotation process, and final saving and downloading of the project).

The workflow of the tool is all seamlessly connected in one window, and is depicted in Figure 3. First, the user can choose either to upload a new video, upload an existing *json* annotation, or select a previously-saved annotation project. In case of a new video, the tool will process the video in order to detect shot boundaries. Once finished, the interface will be fully loaded with the video, the shot timeline, and annotation labels.

After the video has been processed, the user can then perform a number of operations such as modifying the annotation labels, navigating the video through the video player controls or directly on the timeline, and applying or removing labels to each shot. Finally, the annotation file can be downloaded in *json* format, and it is also automatically backed up onto the server. A sample annotation of a shot is formatted as such:

```
{"annotations":[
    {"start_time":0,"end_time":9.08},
    ...
    ...
    {"start_time":24.92,"end_time":26,
     "labels":{
        "Shot size":["MCU"],
        "Scene Type":["action"]
    }},
    {"start_time":26,"end_time":26.84,
     "labels":{
        "shot-size":["medium-closeup"]
    }},
    ...
    "labels":{
        "special":["over-the-shoulder","cowboy","silhouette"],
        "Shot size":["MS","MLS","LS","MCU","CU"],
        "Scene Type":["action","dialogue"]
}]}
```

As an improvement from existing annotation tools, MIDB has integrated shot boundary detection, and no limitation of the video length. The annotation labels can be customized directly using the tool, requiring no programming experience whatsoever, and for advanced users, the javascript framework makes it easy to design and add additional functionality through plugins.

3 Future Work

Our goal is to publish MIDB and collected data on a server for public or limited academic use, and to eventually establish a high quality film dataset that can be used for virtual cinematography, film analysis, and data science. To this end, the next steps to this project involve creating an interface for user-designed plugins such as automatic face detection or shot-type assignment, integrating a selection of public domain films from Internet Archive (`https://archive.org`), and running pilot studies on the tool's ease of use.

Another goal is to establish a set of film practice and theory validated vocabulary, and allow the selection of these vocabulary sets to be the default annotation labels. This is also with regard to the recent development of film analysis and directing languages for virtual cinematography [WC15][RGB13].

References

[AP05] Olivier Aubert and Yannick Prié. Advene: active reading through hypervideo. *Proceedings of ACM Hypertext'05*, pages 235–244, September 2005.

[Kip12] Michael Kipp. Multimedia annotation, querying and analysis in anvil, 09 2012.

[MWS+15] Billal Merabti, hui-yin Wu, Cunka Bassirou Sanokho, Quentin Galvane, Christophe Lino, and Marc Christie. Insight: An annotation tool and format for film analysis. In *Eurographics Workshop on Intelligent Cinematography and Editing*, page 57, Zurich, Switzerland, May 2015. The Eurographics Association.

[RGB13] Remi Ronfard, Vineet Gandhi, and Laurent Boiron. The prose storyboard language: A tool for annotating and directing movies. Chania, Crete, Greece, 05 2013.

[WC15] Hui-Yin Wu and Marc Christie. Stylistic Patterns for Generating Cinematographic Sequences. In *4th Workshop on Intelligent Cinematography and Editing Co-Located w/ Eurographics 2015*, pages 47–53, Zurich, Switzerland, May 2015. Eurographics Association. The definitive version is available at http://diglib.eg.org/.