
CS 4170 S25 Final Project Proposal

Luke McMeans
cet3qt@virginia.edu

Ansh Pathapadu
tcq7wn@virginia.edu

Abstract

Oftentimes, the nuanced beauty of film goes unnoticed. Our artificial-intelligence-driven solution aims to shed light on the details and genius of your favorite directors. Powered by natural language processing, neural networks, and more, our scene analysis tool provides users with an opportunity to hear perspectives and theories about the scenes they find intriguing.

1 Problem Definition and Significance

When you're watching movies, you are immersed in the process. You don't realize the small details behind each movie, scene, and shot. You don't realize why those details mattered or why the filmmaker spent weeks deciding on those specific details. But once you take a class or watch a few videos, these details jump out at you, and you start to pay more attention to the art of filmmaking.

However, the average viewer does not have the time or effort to go and take these classes or watch these videos. They often miss the complex techniques of cinematography, editing, lighting, and sound design that a director uses to craft a simple scene. This then creates a gap between those who study film theory and those who watch movies for fun. To bridge that gap there are potential solutions such as pre-recorded director commentaries or static film analyses, but they are passive experiences and do not adapt to the viewer's needs or the specific scenes being watched.

Many directors and producers have included a commentary viewing alternative, welcoming users into their creative minds during the creative process. However, these were more common in owned, physical media, but have become rare in the age of streaming services [1]. Along with this, it is hard to cover every detail in a commentary viewing. The directors are often too busy to answer all the questions, leaving many of the audience members still curious.

2 Proposed AI Solution

AI can help to bridge the gap by offering a similar type of commentary on the film by analyzing and interpreting cinematic visual and auditory elements in real time and then producing intelligent reactions to a film's artistic decisions. An artificial intelligence-driven tool can harness computer vision, natural language processing, and speech processing to provide feedback about a film dynamically while pausing or altering its narration so as to not interrupt critical dialogue. This technology provides not only a richer viewing experience, but also enhances the interactivity of film education.

Our movie player will have an interactive segmentation tool that allows users to select the portions of a film they would like to have analyzed. Our model would then process through the clips, and generate a commentary. These would be purely theories as our model can't speak for the director's vision, but it still provides a gateway into enhanced knowledge and appreciation for the art form. The model would also process the script at that point to ensure an accurate understanding of the dialogue, character actions, and additional aspects of the scene(s).

The foundation of our visual processing would come from a neural network. Using this, we can dig through layers and features to detect complex patterns and aspects of lighting, cinematography,

blocking, and more. As for the script, the model would utilize Natural Language Processing (NLP) to read and interpret character dialogue/delivery. For a holistic view of the model, we can combine textual and visual elements in multimodal analysis. In 2022, MIT graduate student Alexander Liu was able to recognize auditory and visual inputs and describe what it's observing (for example, "a man juggling and laughing in amusement")[2]. We hope to expand upon the strategies of Liu and members of the Computer Science and Artificial Intelligence Library and apply them to a film context, where we can then generate our commentary. We can build off of this by using a film theory knowledge graph, finding relations to the visual/auditory choices to greater concepts of theory.

When generating our commentary, we need to find out when to present it. To do this, we can use Markov Decision Processes and Utility-Based Decision Making to figure out exactly when to speak as a scene is playing. For this to be functional, our states would be the scenes themselves – certain frames, movements, emotion of dialogue, etc. Our actions would be the moments of commentary based on the orientation of states. Along with this, these processes can also be used when generating the speech of our commentators, using transitional probabilities in language modeling to create/train natural speech patterns and flow. To reinforce our model, we can add a reward system to train the model better, having users rate their satisfaction with the commentary and whether they felt it contributed to the understanding of the scene(s). Our hope is to create an ideal policy function $\pi(s|a)$ for each user, ensuring that we are reasonable in our interpretations and provide relevant details to the desires of our users.

With this tool, we hope to expand access to film education and give users a new perspective on the films they love. Powered by artificial intelligence, personalized solutions tailor to their desires and ensure that their questions are answered.

References

- [1] Larman, A. (2022) The death of DVD commentary: How the rise of streaming platforms has killed home entertainment. *The Critic*. <https://thecritic.co.uk/the-death-of-the-dvd-commentary/>, Accessed: April 4, 2025.
- [2] Zewe, A. (2022) Artificial intelligence system learns concepts shared across video, audio, and text. *MIT News*. <https://news.mit.edu/2022/ai-video-audio-text-connections-0504>, Accessed: April 4, 2025.