

Spring 2023 CSCI 576 Multimedia Project

Instructor: Parag Havaladar

Demo date: Wed May 3rd, Thu 4th, Fri 5th 2023

The course project is meant to give you an in depth understanding of some of the areas in multimedia technology. Since this is a broad field, there can be a variety of interesting projects that can be done depending on your interests which can also extend to related and complementary topics that are taught in class.

Also, I have often found that a larger project can be successfully accomplished via collaboration. Additionally, working together to design and integrate code can be a rewarding exercise and you will frequently need to work in teams when you set out to work in the industry. *Accordingly, please form groups of three, and utmost four students.* We have started a discussion board to help you make groups, where you may post your preferred language of implementation, availability etc. Once your group is decided, please send the TAs an email so we can assign a time for demo on the due date. Remote DEN students may form groups with in-class students. If you are a remote student and have trouble finding a partner, please send an email to the TAs and we will try to facilitate group formation. The demonstrations will be done online over zoom. You will be asked to submit code in certain cases for further evaluation.

This semester, we are proposing a project to extract structure from media elements like video/audio with indexes into the structure and alongside build an interactive media player that will allow you to explore the media.

Extracting indexes from video along with interactive exploration

When you open a book, you normally have an index that has been created by the author, storyteller, writer etc. The index gives you a good description of how the book is organized with pointers to page numbers to the various chapters. In an online setup, this index can be hyperlinked to different documents or within the document allowing interactive exploration. For more complex media such as video this ability of providing an index with and interactive exploration is also possible but currently limited to well authored or pre-authored content eg a DVD movie, or BlueRay movie, or when watching a movie on a platform like Netflix, where episodes can be selected. But such a setup is not directly available unless it is created, mostly manually, to then be consumed interactively.

In the framework of video analysis for indexing application (for example application related to MPEG-7), the representation of video sequences is an important issue. It is not enough to describe the content of a video but also to develop techniques which are able to automatically create these descriptions. These descriptions are useful for interactive exploration when they have an indexed mapping into the video sequence, which is also generated as a part of the process. Your solution should rely on the visual and auditory information in the input video file. The analysis learnt in class lectures provide you a framework to implement a solution, and you are also welcome to use external libraries, research etc.

Specifically, in this project, you are asked to develop a solution to

1. Arrive at this logical index or table of contents in an automated manner given a video with audio as input.
2. Once the index or table of contents is extracted, you want to show it in an interactive player setup where you can jump to any index to browse the video/audio and change it interactively allowing interactive exploration.

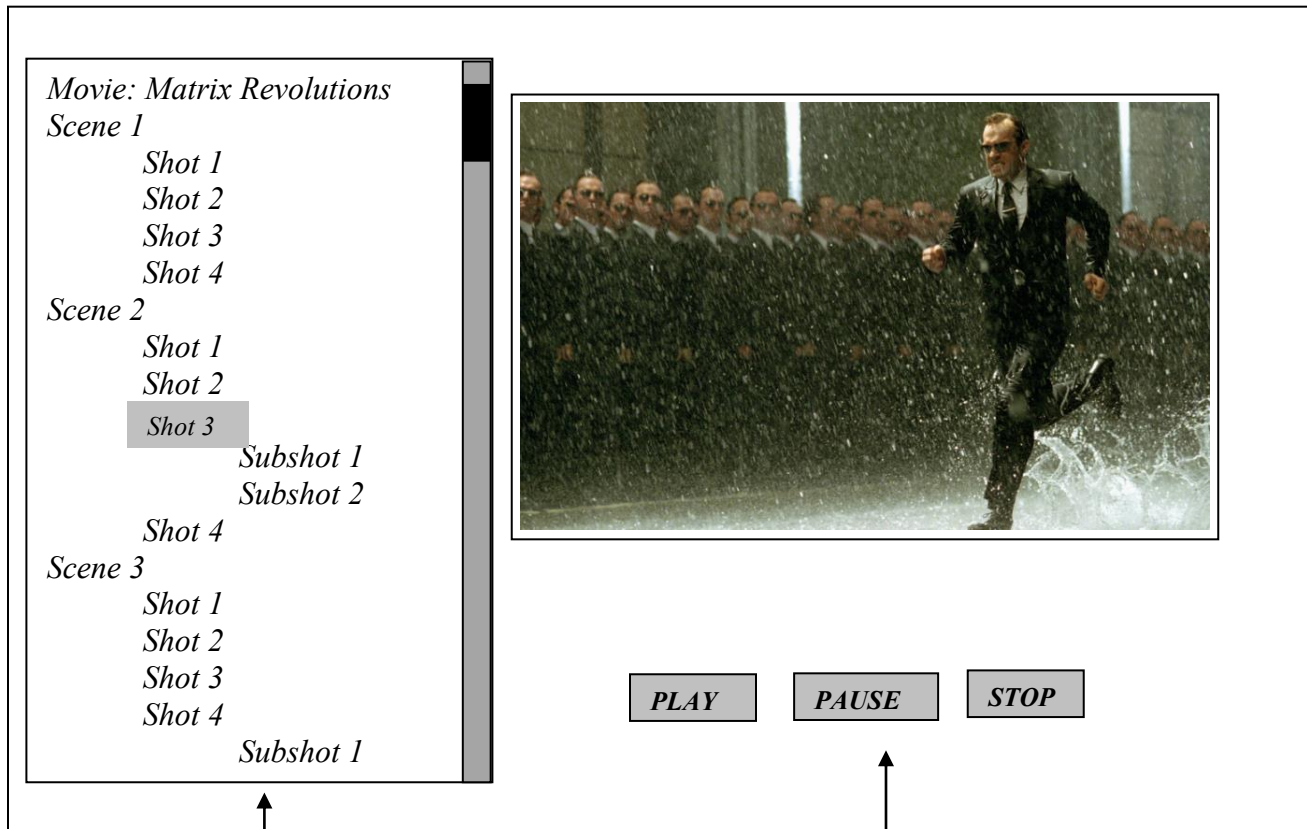
Your project will be called as follows: “*MyProject.exe InputVideo.rgb InputAudio.wav*”
The InputVideo.rgb file and InputAudio.wav file are synchronized. Furthermore, the video width, height and fps will be fixed, also the sampling rate of the audio will be fixed. But the length of the video can vary.

Evaluation of your project will be influenced by a few guidelines noted below:

- Correctness of your table of content given a new input. This should be as hierarchical as possible. Look at the example illustration in the interactive player description.
- The interactive exploration of your player. Our expectation is that bouncing around indexes updates the player interactively and in a synchronized (audio video) manner.

Interactive A/V Player and Interface

Design a simple *functional* interface that shows the extracted table of contents listed along with the video playing on the right. You should have support for play pause and stop.



Interface showing video structure.
Scene2, Shot 3 is currently playing

Interactive buttons
to play, pause, stop

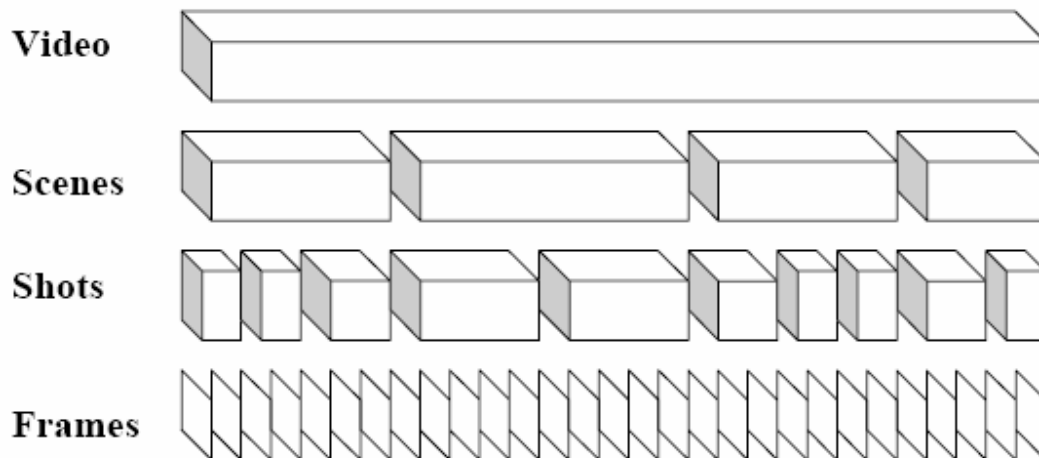
Expectations and Evaluations:

We do expect that you analyze the problem space and produce a solution. The answers may be subjective, but we expect unambiguous hierarchies on clear datasets. When it comes to playing audio and video, we do expect them to be synchronized. Additional expectations:

- You should be able to play, pause, stop the video (and its corresponding audio). Pausing and playing resumes the video from the paused time. Stopping and playing restarts the video from the beginning of the current selected shot.
- There should be an area where a hierarchical video table of contents is visible. This is the output of your processing. There is no need to use a fancy UI (like a tree hierarchy) but the shots should all be displayed (via a scroll bar if needed) and elements should be “selectable”. The scene/shot don’t need contextual names, a number index should be sufficient – as shown in the figure above (left).

- The hierarchical table of content should show a breakdown of the video at various levels:
 1. Sequence/scene level
 2. Individual shot level
 3. and if the shot is varied properties, then at the subshot level.
- The anatomy of a video figure below might help with this. In a movie context, a scene (sequence) indicates a group of shots with a context eg, all shots in the “apartment” room, or all shots in the “garden” scene. All the shots in a scene (sequence) are continuous and have common elements – background environment, actors. When a scene changes, the background environment, properties of sound levels all change. Within a scene, shots are demarcated by an abrupt discontinuous change. If the shot has varying properties of interest (motion, noise levels), then the shot should be broken into subshots. For example, a shot may consists of times when the motion or slow or no motion, and times when the motion is high - these could serve as separate subshots.
- Interactively selecting a segment should highlight the segment. If the player is stopped, the upon playing the video should start playing from the beginning of that segment. If the player is playing a different segment is selected, the player should update by playing the new selected segment. This way you can bounce around the content by selecting various shots
- As the video plays continuously the highlighted segment selection (scene, sequence, shot) should correspondingly update.

The anatomy of a video might help better understand how to build hierarchies.



- **Frame:** a single still image from a video, eg NTSC - 30 frames/second, film – 24 frames/second
- **Shot:** sequence of frames recorded in a single camera operation
- **Sequence or Scenes:** collection of shots forming a semantic unit which conceptually may be shot at a single time and place