# Spring 2021 CSCI 576 Multimedia Project

# Instructor: Parag Havaldar

# Projects due by: Thu May 6th 2021 & Fri May 7th 2021

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 large 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 *two/three 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 fill the form sent by the TA via email so we can assign a time for demo on the due date. If you are a remote student and are having trouble finding a partner, please send an email to the TA/me and we will try to help. Also, if you are a remote student, we normally allow you to do a remote demonstration. Details on this will be decided soon.

This semester, I have proposed a project in the area of quickly exploring a large directory of media information. The motivation, description and the expectation for the project follows on the next page.

# Video Summarization

With the advances in inexpensive digital video and sound recording instruments, video /audio data is getting to be common place now. There is a lot of digital video information everywhere – whether it be personal recordings of event or whether it be high end content streamed to you via the internet and cable networks. With a lot of video information, there has to be a way to browse through it efficiently and conveniently. This is true for other simple media types such as text and to some extent images. For instance, if you load a big text document, and want to be able to quickly read about information on a specific topic, you can easily search through the entire text using a “string”, and read relevant topics. Or better yet, have an automated way to programmatically going through an entire text document and extracting pages that may be of relevance for you to sift through quickly and take necessary action. This is not straight forward for a media type such as a video/audio.

The motivating question here is how you quickly browse a video in order to extract the most useful information that the video has to offer. Here are a few example cases

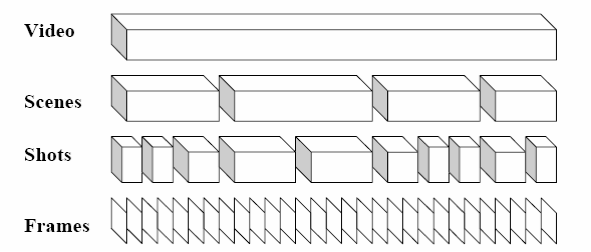
* You want to watch a video recording of a sports game, which is two hours long, but you want to do this in a few minutes, or just see the highlights of the game.
* You want to quickly browse a surveillance video to see interesting events that might have happened.
* You want to watch a movie ninety-minute feature in ten mins.
* You want to go through a documentary or a video interview of a person but only for topics that are relevant to you.

In this project, you will develop a methodology to “summarize” videos by abstracting the main or important occurrences, scenes, etc. to provide an easily interpretable or a highlight synopsis of the video. From a user’s perspective, this is a more viable way of quickly understanding/evaluate a large video. It is not easy to extract the semantics of what “important” means in a video and is also often subjective. In the project, you will base your analysis on different qualitative ideas learnt in the class.

Here is a list of concrete tasks that your video summarization project needs to achieve,

1. Create a video/audio player that will display them synchronized.
2. Break the input video into a list of logical segments – shots (see anatomy of a video below)
3. Give each shot a quantitative weighting value depending on the motion characteristics (eg fast motion is more important than slow motion), audio characteristics (eg high audio levels, such as cheering, are more exciting than low audio levels or silence), color characteristics (eg a variance of colors can be construed to be more important than constant colors) – or any other .
4. Using the above characteristics, piece together a new video using specific high weight age shots or even parts of shots.
5. Display the video/audio.

Anatomy of a video:



* **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