CSCI 576 Multimedia Project

Instructor: Parag Havaldar

Demo date: Wednesday May 4st & Thursday May 5nd 2022

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 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 **at least two but utmost three** students. We have started a discussion board to help you make groups, where you may post your preferred language of implementation, availability etc.

This time I want to suggest a topic that involves removing specific, sometimes unwanted content from audio video data. On the next page is a complete description giving the motivation and the requirements of your project. It should be a comprehensive project that aims to increase your fundamental understanding of rendering video and audio in a synchronized manner, and analyzing the signals with topics discussed in class.

Detecting and Replacing Advertisements in Multimedia Content based on Brand Images/Logos.

There is an increased amount of video and audio content broadcast and streamed everywhere today. Such content needs to be frequently analyzed for a variety of reasons and applications – such as searching, indexing, summarizing, etc. One general area is in modifying content is to *remove/replace* specific parts of frames or even a number of frames altogether. Let's consider removal of frames from video & audio content. The motivating question that you will need to think about in this project may be described by - how do you *automatically* analyze video/audio and remove frames that correspond to a specific description. General examples of this type may include:

- You want to watch a video recording of a sports game, but you want to remove all the non-interesting areas and see only the sections that have good plays and goals scored.
- On the audio side, you must have seen "bleep censoring" or "bleeping" which is defined as the replacement of profane words, or even classified information with a beep sound – usually a 1000Hz tone
- You want to quickly process a long, mostly boring surveillance video and cut out the uninteresting parts so that only desirable sections of "interesting events" can be highlighted.
- You want to remove all video frames from a recorded video that shows a specific person or a copyrighted object in there.

As you might guess, the gamut of this problem space is very vast and hard depending on how well you can describe the semantics of the content that need to be removed. While the general-purpose problem of describing "unwanted" content is vague and thereby difficult, we can certainly make the problem space easier by defining exactly what you want to remove and the datasets on which you want to operate. One such specific application area is analyzing advertisements.

Advertising is a source of revenue for content owners and content distributors but often proves to be an unwelcome viewing hindrance when the intended audience is consuming the content. Furthermore, if the advertisement has to be effective as a marketing tool, it is more useful to have advertisements targeted toward specific consumers rather than a class of consumers. In other words, although you and I watch the same streaming video, we should see different advertisements depending on our specific likes and interests. For live or on-time broadcasts, removing/replacing advertisements is not possible since the content is linearly delivered. However, given the proliferation of inexpensive digital video recording devices (DVRs) which now come integrated with your setup top or cable boxes for television content, it is much easier to record and watch video in post. Also, when you pay for streaming content, you are able to download it for yourself on a variety of open and proprietary platforms such as your laptops, tablets, kindles etc. In such

cases, by preprocessing the video, you should be able to remove advertisements and optionally even replace the advertisements with something more targeted.

The next question is how you define the semantics for replacement. One practical choice might be (and which we do not have) based on the browsing habits of the individual. Another choice (which we have in this case) might be advertisements based on the video itself e.g. if there is a coke symbol, or a star bucks symbol observed in the video, you might want to insert a corresponding ad by that company.

In our case let us define the problem as follows –

Design an algorithm to automatically remove advertisements from the video (and corresponding audio) which is interspersed with advertisements Furthermore, extend this process to detect the a given specific brand in the video and if present, replace the original advertisement with a corresponding topical advertisement.

As input, you will be given
- Video (and corresponding audio) files with advertisements in them. You may assume that the advertisements have smaller time segments compared to the videos they are embedded in and different audio characteristics

Note all input data files will be in the form of image frames where each frame size, fps are the same. Also their corresponding wav files will also have the same sampling rate. The advertisements in the video might be at different temporal location

- Brand image files as shown above which you will need to detect in the input videos
- Brand advertisements in the same format(.rgb + .wav) that you will need to replace in the video

Correspondingly you are supposed to devise two programs:
1. You will also need to create an audio video player which will be run as
   MyPlayer video.rgb audio.wav
   Your player should be able to synchronize the video and audio rendering based on the video frame rate and the audio sampling rate. You should devise a simple user interface to **play**, **pause** and/or **stop** video.
2. You will need to create a program that takes as input a video/audio stream with advertisement sections and creates a new video/audio stream with the advertisements removed.
   MyProgram inputVideo.rgb inputAudio.wav outputVideo.rgb outputAudio.wav
3. You will need to enhance part 2 (with optional arguments as necessary) to analyze the video frames to see if any known brands exist. If a known brand is detected, then create a rough outline/bbox in the frames that it appears, replace the following next advertisement with the given brand advertisement. Note – your insertion/replacement process should insert video frames and audio data at the right place so that when played everything seems synchronized appropriately.

Here are some guidelines to help you design your project. You have to understand and define the characteristics which are common to advertisements – what makes a group of frames an advertisement? Some common characteristics which might serve as heuristics may the presence of one or more of the following -
- Advertisement sections are shorter and rapidly changing
- Audio levels change suddenly from the main correlated section.
- Fast motion in sections uncorrelated with the main content
- For detecting whether a brand image is present in a frame, you can use color space analysis, where you try to match the colors in the brand image to the colors in a frame.

These are ideas that you can implement based on what we have learnt in class or extensions thereof, but you are welcome to research and use different approaches.

Here is a list to give you an idea of concrete tasks that your project needs to achieve:
1. Read in the input video/audio – remember you might not be able to fit the entire content in memory for processing.
2. Break the input video into a list of logical segments – shots (see anatomy of a video below) How can you achieve this?
3. Give each shot a variety of quantitative weights such as – length of shot, motion characteristics in the shot, audio levels, color statistics etc.
4. Using the above characteristics, decide whether a shot or a group of adjacent shots might be an advertisement
5. Remove the shots that correspond to the advertisement. Write out the new video/audio file.
6. If brands are detected, replace the old advertisement with a new advertisement to write out the new video/audio file.

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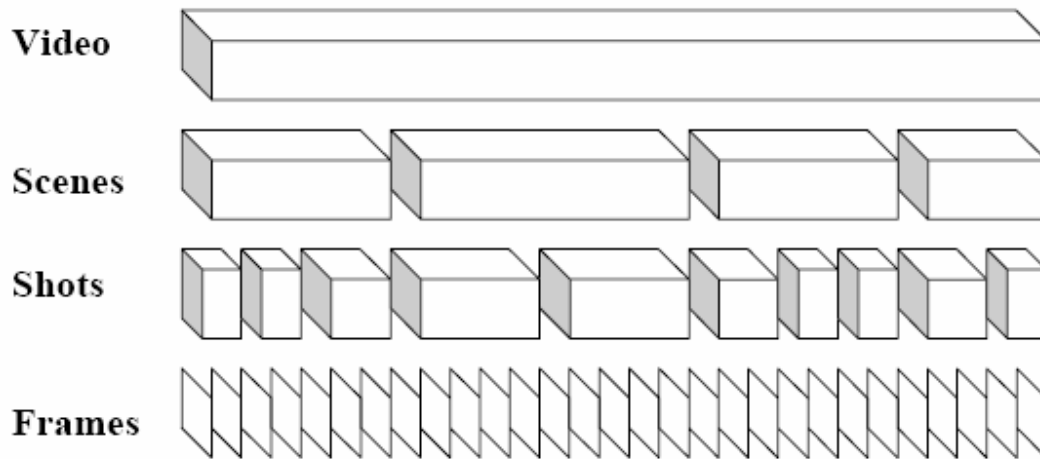
- **Frame**: a single still image from a video, eg NTSC - 30 frames/second, film – 24 frames/second
- **Shot**: sequence of frames recorded in a single camera operation
- **Sequence or Scenes**: collection of shots forming a semantic unit which conceptually may be shot at a single time and place

The evaluation and grade will be based on a variety of tests including – video audio synchronization, detection and remove of advertisements, detection of brand logos as well as appropriate and synchronized replacement.

NOTE: This is a hard problem to solve in its entire and general scope, but for the project, we have limited the scope and given well defined datasets on which your algorithms should work. The video/audio synchronization is a straightforward implementation to evaluate but the detection of advertisements and even more so, the detection of brand images may have dubious answers. The answers you arrive at might not be wrong algorithmically, so please make an effort to display results appropriately to help us evaluate your algorithms – eg in your output videos, you should draw rectangles around the detection areas (whether right or wrong) so that we can analyze how your algorithm worked.