Fall 2024 CSCI 576 Multimedia Project

Instructor: Parag Havaldar

Demo date: Wed Dec 11th, Thu Dec 12th, Fri Dec13th 2024

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.

We 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 two, and utmost three students.* Please make use of Piazza threads to communicate preferences and 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 register you as a group. 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 that builds upon your assignment 3 compression on images and applies it to video, in a more structured manner to qualitatively improve foreground compression vs background compression.

Foreground/Background segmented compression

This project is an interesting application of block-based motion detection and compression in video-based communication industries including entertainment, security, defense etc. Here you are required to separate an input video into different layers based on motion characteristics - those that represent the foreground layers and the background layer. Each layer is then compressed differently according to input parameters provided. The theory taught in class dealt with computing motion vectors to help better compression of video, here we use it to segment a moving region and then use the same for compression.

To formally describe the project –
Given video frames as input, process the video to find out objects in motion or macro blocks that compose objects in motion and macro blocks that comprise the background. Once these macroblocks are detected, each one is compressed using quantization parameters provided as input. You will be writing two programs that logically share the same functionality – a layer based encoder, that will take as input a video file and two quantization parameters to produce a compressed file and a decoder that will take as input the compressed file to and generate a rendering of the final output. The specifics of the input arguments to your two applications are described below.

1) myencoder.exe input_video.rgb n1 n2
input_video.rgb is the input file to your encoder.
n1 represents quantization step for foreground macroblocks
n2 represents quantization step for background macroblocks
the output of this is a "input_video.cmp" layer wise compressed video file
2) mydecoder.exe input_video.cmp input_audio.wav
decodes and renders the compressed video file in a group designed A/V player with synchronized audio-video playback. The player should have the capability to stop/pause/play/step to observe frame by frame results.

The whole project and the applications may be divided into four main parts described below.

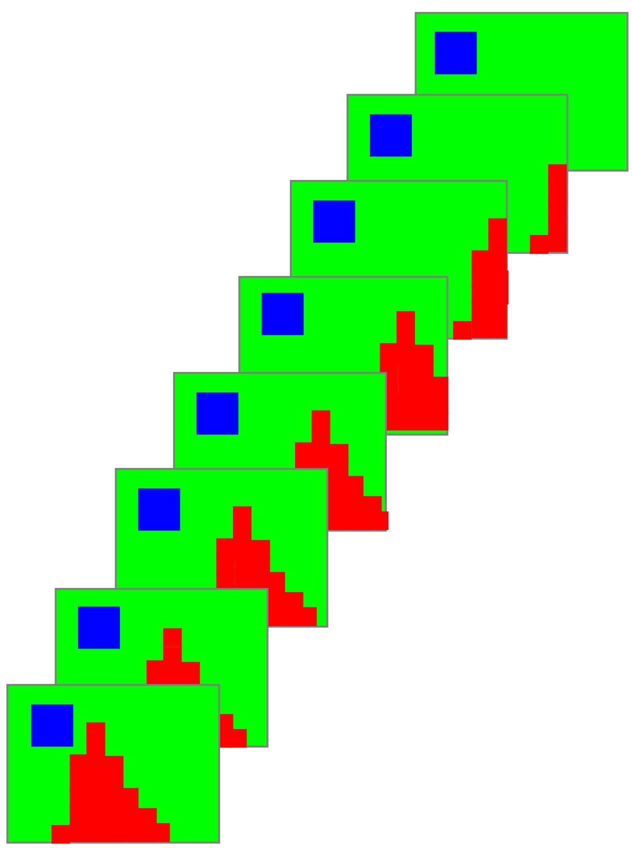
1. Video Segmentation (in myEncoder.exe)
You are required to divide each frame of the video into **background** and **foreground** macroblocks and further group the macroblocks into separate contiguous groups, each of which approximates either a foreground object or the background . Note that your macroblocks may not be precise, especially at the boundaries but nevertheless you should be able to segment out contiguous blocks of foreground elements and background. These are the guidelines you should follow for this part:

1. The first frame may be assumed to be an I frame, but for every successive frame divide the image frame into blocks of size 16x16 pixels
2. Compute the Motion Vectors based on the previous frame – this is like the block-

based MAD (mean absolute difference brute force search) or even better any fast motion estimation (FME) technique. At the end of this step, each macroblock of the frame should have a motion vector.

3. Organize the blocks into background and foreground based on the similarity of the directions of their motion vectors. Background macro blocks either have a close to zero motion vectors (if camera is not moving) or a constant same motion vector (when the camera is moving). On the other hand, foreground macroblocks have similar motion vectors with macroblocks of a moving region normally being connected or contiguous. Thus, the main criteria for grouping regions is
   a) Contiguous or adjacent
   b) The motion vectors are all consistent – *important!* The consistency of the motion vector direction gives you an indication that all the macroblocks probably belong to the same object and are moving in a certain direction.

Note that you can have different regions moving in your video sequence. By the end of this first part, you should know for each frame for each macroblock – whether they represent a foreground or a background macroblock. The general-purpose solution of this problem is not easy, but as long as your program comes up with a good demarcation of foreground and background based on motion, that is fine. An example is demarcation is shown below where the green macroblocks represent the background, and the red/blue macroblocks represent the foreground elements.



2. Compression (in myEncoder.exe)

After segmentation and a determination of which macroblocks are foreground/background you are required to compress the background and foreground macro blocks using different quantization values. The quantization used for quantizing the DCT coefficients of each block is controlled by the input parameters *n1* and *n2*. Background macroblocks will generally be more quantized whereas foreground macroblocks will have lesser quantization. The requirements here are

1. Divide each frame into 8x8 blocks (similar to MPEG). Each block is either part of a foreground macroblock or a background macroblock.
2. Perform a 8x8 DCT, quantize the DC/AC coefficients using a uniform quantization function where each number in the quantization table is 2^n where the exponent n will be $n1$ or $n2$.
3. Scan frame 8x8 blocks in scan line order and save the quantized coefficients of each macroblock into a compressed file. This file will be used by the decoder to decode and render the results of your compression. While you are free to choose any file format for your compressed file, here is a simple suggested format that you can make use of.

Fore a given input file – input_file.rbg – produce a compressed file input_file.cmp that contains the following -

n1 n2
block_type coeff1 coeff2 ... coeff64 coeff1 coeff2 ... coeff64 coeff1 coeff2 ... coeff64
block_type coeff1 coeff2 ... coeff64 coeff1 coeff2 ... coeff64 coeff1 coeff2 ... coeff64
block_type coeff1 coeff2 ... coeff64 coeff1 coeff2 ... coeff64 coeff1 coeff2 ... coeff64
block_type coeff1 coeff2 ... coeff64 coeff1 coeff2 ... coeff64 coeff1 coeff2 ... coeff64
block_type coeff1 coeff2 ... coeff64 coeff1 coeff2 ... coeff64 coeff1 coeff2 ... coeff64
....
...

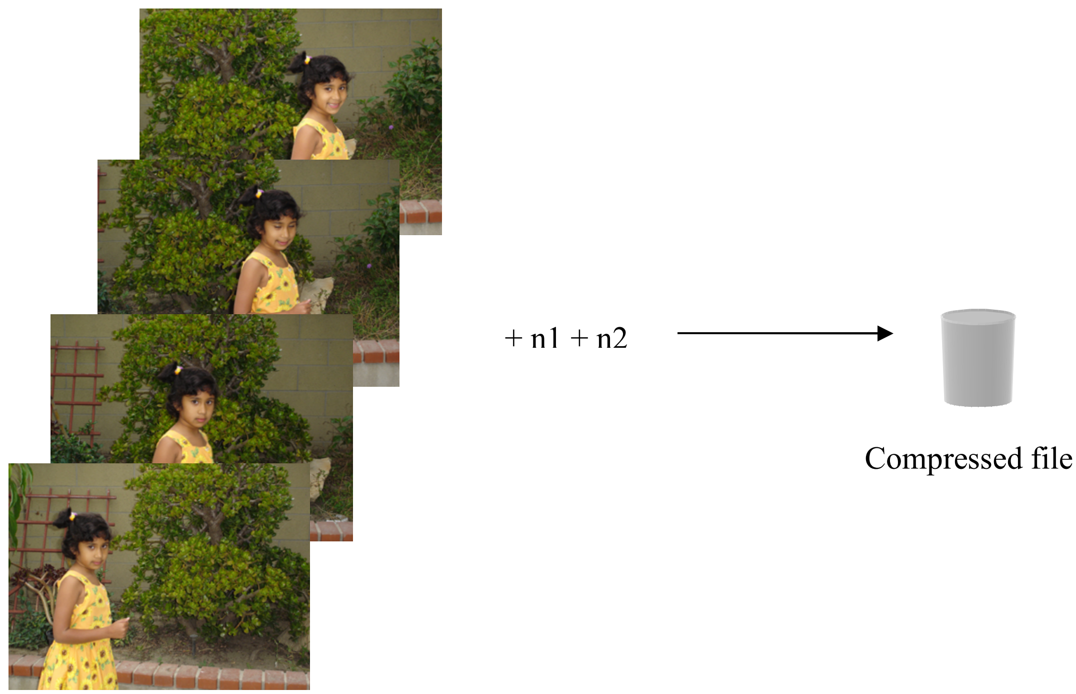
Where n1, n2 are the quantization inputs, block_type is a binary state to say whether each 8x8 block is a foreground or background block and the coefficients are the quantized DCT coefficients for each block. Note that for each block there are three sets of sixty-four coefficients, a set for each of the R, G and B components. **Note here you are NOT expected to perform zig zag ordering, intermediary representations and entropy coding as per the JPEG standard**.

3. Decompression and Decoding (in myDecoder.exe)

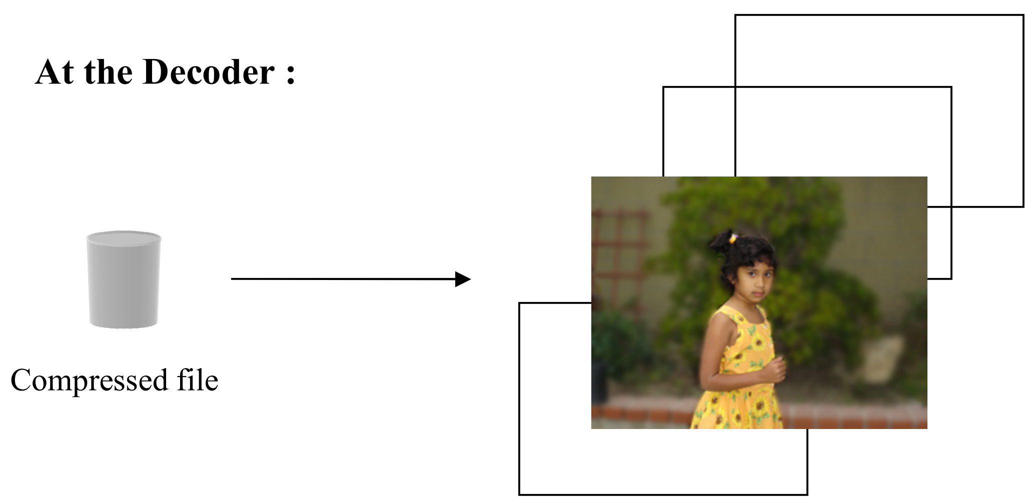
The decoder will take the input_file.cmp as input. You should be able to parse the file to get the n1, n2 values and the coefficients of each macro block. Additionally, you should be able to dequantize the blocks using either *n1* or *n2* depending on whether the block_type represents a foreground object or a background object. Decode and display the results of your decompression.

Here is an example workflow with qualitative results. Note that in the decoder the foreground macroblocks (macroblocks in motion) have less compression artifacts than the background macroblocks. Note the effect would have been reversed if the values of *n1* and *n2* were interchanged.

At the Encoder:



At the Decoder :



A few implementations notes

- You will have to create your own methods to display the input and output videos. This will include playing the video at frame rate synchronized with sound, with an ability to play, pause and step through the video frames so that we can evaluate your compressed-decompressed output to the original. The UI of the player is not important as long as the required functionality is supported.The input datasets and a description will be kept on the class websites and the formats should remain constant. You may assume that your video width and height are each multiple of 16 and the frame rate is fixed as described on the class project site.
- We will only test the quality of your output and the synchronization of your audio-video playback at frame rate. Correspondingly, we are not going to keep any marks for your implementation and runtimes to compute results – as long as you are able to run and demonstrate your results in your demo slot. A good estimate to hit is end your encoding and decoding process within 10 mins for video lengths given.
- Remember to do macroblock searches only using data of one component (not all three!). Typically, you could convert the image format to YUV and use the Y component for all evaluations. If not, you may just use the R component. However, you will have to do a DCT compression/decompression on all the three channels.