Spring 2025 CSCI 576 Multimedia Project

Instructor: Parag Havaldar

Demo date: Wed May 7th, Thu May 8th, Fri May 9th, 2025

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.

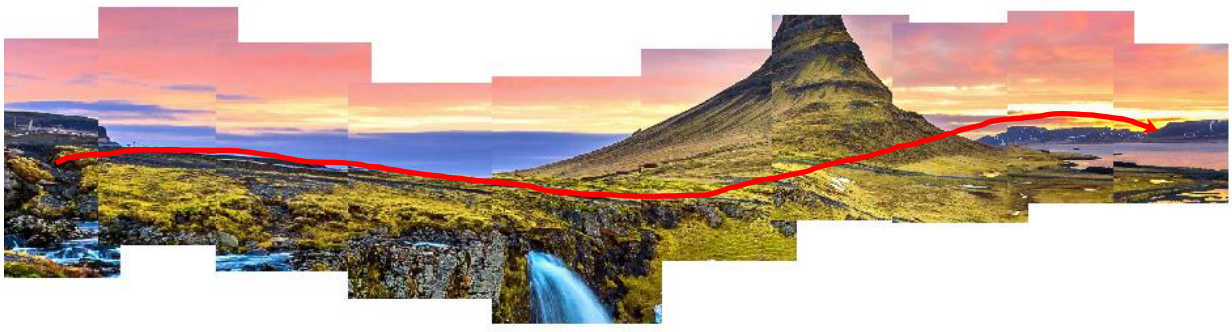
Also, larger projects can be successfully accomplished via collaboration with multiple students working in a group. Additionally, working together to design and integrate code can be a rewarding exercise and you will frequently need to work in teams when you graduate and work in the industry. *Accordingly, please form groups of **at least two**, but **utmost three** students.* If you need help forming groups, please use Piazza discussions, where you may post your preferred language of implementation, hours of availability etc. Once your group is decided, please send the TAs an email so we with the organization around the project and project grading. Remote DEN students may form groups with in-class students. If you are having 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 where all members must be present and on camera for Q&A.* You will be asked to submit code in certain cases for further evaluation.

This semester, we are proposing a relevant project to search and index into video/audio with interfaces to show your results in an interactive media player. Details are explained on the following pages. For this project, we are placing no restrictions around the language of use or libraries to use. You are welcome to use external libraries, environments of your choice – as along as you are able to fulfil the objective of the project evaluation.

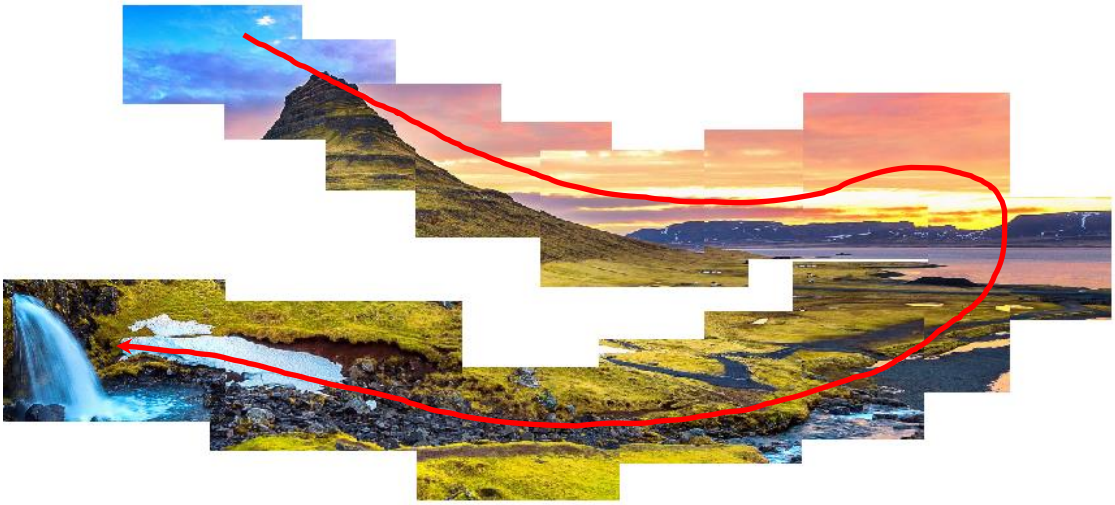
Large-Scale Ultra-High-Resolution Images from Videos

With the advances in inexpensive digital image, video (and audio) capture devices, media data is getting to be commonplace now. There is a lot of digital video information that people can capture with their phones, consumer cameras and other similar devices. The convenience of capture is no doubt a big advantage with such setups, but a worthwhile drawback to think of is that all these images have limited focal view, and as such have a “smaller” window into the whole scene that you want to capture. You can change the focal length to be a wide zoom to get a wider view, but you still capture an image at the same limited resolution setting that the camera offers. To capture a “larger” window of the scene requires a more expensive and calibrated setup that is probably not possible at consumer levels today. Take a large ultra-high-resolution image (think Gigapixels) - this either requires expensive setups or custom software calibrated setups. It should be possible to create such large-scale images by putting together limited but continuous framing content from a video.

When combining images to create such high-resolution content, a well-known paradigm that is normally followed is to take pictures with your device involving a variety of angles and ensuring a good overlap. These images can then be stacked and stitched together to create the final large image. On the one hand, this process is easier to manage because there are a few static images, but getting overlapping content needs thoughtful attention and calibration and the need for precision can be cumbersome. Another easier workflow might be using your phone/video camera like a paint brush where you zoom into a scene and “paint” your scene to capture a video. A few paint strokes or a few videos can quickly and effortlessly capture details of your scene. The video scans below show such examples. A computational process then should be able to sift through all the frames and put them together to create your ultra-high-resolution image.



Video scan example with the red line showing direction of capture



More free flow examples of capture with the red line showing direction of capture

All the scans can then be put together seamlessly to create a large panorama as shown below. To formalize the problem – *in this project, you are asked to take multiple videos of the scene and output a ultra-high-resolution good quality image.*

Displaying the output:
You may store the out as a normal image which can be uploaded in a photo viewer and explored for quality. We want to see the high-resolution image and be able to zoom in on sections to see how your algorithm performed when stitching images together.

Processing the Videos:
Please see class discussions on how to find correspondences that can be used to compute image transformations from translation, rotation, scale and homography that will help align images.

Expectations and Evaluations:

We do expect that you analyze the problem space and produce a solution. The final results are not subjective since we know what the output will look like. The evaluation will include:

- The correctness of your image including all the coverage content
- The quality of your stitched high resolution image including exploration of how seamlessly the assembly of images appear.

- The time your algorithm takes to process and create the high resolution output from the input videos
- Oral Q&A about your algorithms and process.