Computer Vision

Semester Project Part 1 Deliverable

Daniel McCarthy and Chris Hunt

(Actual picture dataset collection was performed by Chris while in quarantine)

As discussed previously, our initial dataset that we collected personally is of a keyboard with colorful stickers marking its four corners. The goal of using this dataset will be to train our program to recognize where the piano is in a given image and then perform a series of transformations on the image to straighten it and distort it if need be so the resulting image can be fed more easily into a model that will recognize what notes are being played as if the picture was taken from directly overhead.

To perform the data capture, the camera was positioned above the piano and pictures were taken in 6 iPhone bursts of about 50 photos each. Within each burst, the camera was moved around slightly while still atop the piano. This was done in order to capture different angles without making the piano completely unrecognizable. Between each burst, lighting conditions were adjusted slightly. This was accomplished by changing the blinds on the window between iterations. The three blind positions were down, center, and up, with center letting in the most light. These blind positions were done twice with two different colors emitting from surrounding LED lights: one with a slight green hue and one with a slight blue hue.

It should be noted that all of these conditions used for the data collection did not drastically change how the photos looked. The photos all have an identical piano setup - just ever so slight modifications to the piano’s surroundings.

One effect that was not intended was the resolution changing. In some photos, the quality dropped because it was taken while the camera was moving. This will also provide miniscule changes in the data without compromising recognition of the piano.

Once we collected all 300 pictures forming our dataset, we performed a 60/20/20 split on the dataset. We created 3 folders corresponding to the training, validation, and test data. Careful steps were taken to ensure that each of these folders contained a variety of image types from our original 300 pictures.

Our proposed solution is as follows: The images captured by our video feed will be sent first to a function which will first detect where in the image the piano is. This is made easier by colorful stickers marking the four corners of the piano. Next, the program will decide if the image is distorted in any way from the angle of the camera to the piano and will adjust the image with transformations so that it will now appear as though the image was taken directly overhead. Next, note recognition will occur. The program will need to be able to segment the piano in such a way that each individual note is recognizable and detectable. As of now we are unsure as to how to make this happen. Finally, we need a reliable way to detect which notes are actively being played. A simple solution we have considered would be to detect when a key is occluded by a skin colored object. This could be sufficient to tell if a key is being covered by someone’s finger. However, when we play the piano our hands are often covering many keys without playing them, so this solution is trivial and would only work if a person was playing every key their fingers were covering. This would be our backup solution if we cannot find a better way to differentiate keys that are being played from keys that have fingers resting on them.