

# Data Summary

## Mormon Heritage

Christopher Mertin

From the Tinder API, I have downloaded tens of thousands of photos from people with Mormon Heritage based on their profile on Tinder. To put the value in perspective, I have downloaded more than 2,000 user profiles, where each profile has 5 photos. The resolution of each photo is  $650 \times 650$ .

However, Tinder has recently changed their API settings which makes it difficult to get data from other regions where you are not located so I am attempting to find a work around. If this is not the case, then the work around that I will use is to “clean the data” that I have and separate it into those not from Utah and those with Mormon Heritage.

These images were downloaded as `jpeg` files. From here, the features can be extracted. For example, OpenCV can be used to extract 126 facial features from *each* photo, which can be stored as a numpy array to perform utilities such as clustering. OpenCV will perform affine transformations on the features such that the features are all facing forward. Also, the images themselves can be stored as a matrix so that techniques such as EigenFaces can be used as well. This will be done *after* performing affine transformations on the images so that they will all be aligned.

The data is not yet processed, as it’s still in the raw format, but what I plan on doing is first converting the faces to greyscale, and then cropping the faces to be the same size/area. This is important to get an accurate clustering and EigenFace model.

Simulating similar data is a difficult task as it is photographic data from actual people. Potentially, one thing that could be done is to merge all of the faces together, as with EigenFaces, average them, and then use the OpenCV feature extraction. Then random (but plausible) variations of the features could be made to create new “faces” even though it would just be representing features.