ISyE 6416 – Basic Statistical Methods - Spring 2016
Bonus Project: "Big" Data Analytics
Proposal

Team Member Names: Hugo Acuna, James Netter, Mahadevan Vaidyanathan

Project Title: Facial Recognition

Problem Statement

With a rise in the importance of facial recognition, for not only law enforcement but for websites like Facebook, the ability to create software that has the ability to recognize a human face is paramount. In this project, our goal is to create an effective algorithm for detecting key facial features in a picture of a person's face.

Data Source

Using data from an online source, https://www.kaggle.com/c/facial-keypoints-detection/data, we will test our methods. The dataset contains two-dimensional 8-bit grayscale facial images with corresponding facial points listed in typical (x,y) coordinate format. Each image is represented by a 96 x 96 pixel matrix, whose entries are integers from 0 to 255. These integral values are meant to characterize the intensity of each pixel, for a total of 9216 pixels in each image. The dataset is a matrix of size 7049 x 31, where each row is made up of one facial image, out of 7049, with 30 columns made up of the 15 facial points' (x,y) coordinates, and finally the 9216 numbers representing the pixel matrix of each image melted row by row. The 15 facial points marked in each image correspond to nose, eye, mouth, and eyebrow locations on the face. In the data, though, some images have missing (x,y) facial points, which we will attempt to account for in our final solution.

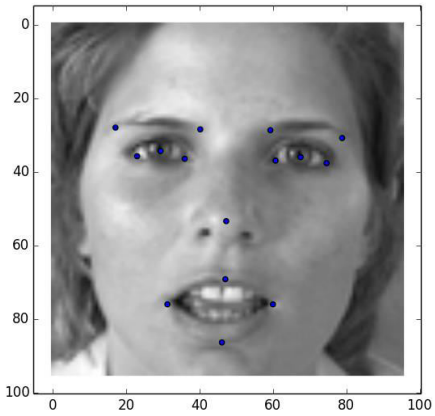

Image 1. Face from dataset, with 15 facial points marked for demonstration

Methodology

Ultimately, with the analysis of the given dataset, our algorithm should be able to identify the facial points of any similar image. In order to do this, we will have to perform several different types of analysis on the given data points. To start, we will consider every face as a vector, consisting of its' 15 facial data points, so we can search for the eigenvectors expressed in each covariance matrices. These eigenvectors can be used to determine distances between facial attributes, making the process of identification much simpler.

To accomplish this laid-out methodology, we will use several analytical methods: PCA, SVD, LDA, and LBP. By using principal component analysis, (PCA), we can take the possibly correlated variables (the facial features) and through dimension reduction, gather more information. Unfortunately, PCA may not analyze the most discriminating features, so we will attempt to offset this with linear discriminant analysis, (LDA). It places more emphasis on the differences from each individual image to find an effective algorithm. We will also analysis the data with singular value decomposition, (SVD), which likely won't compromise the dataset when we analyze it.

By using Local Binary Pattern analysis, (LBP), we will be able to extract local features by converting each pixel into a binary value based on a center point. This will be useful to our efforts because the additional texture classification will reduce the sensitivity of our feature detections to illumination of the picture. To accomplish this benefit, we will take the middle pixel and set this as a threshold value, and each surrounding pixel will be changed based on the middle value.

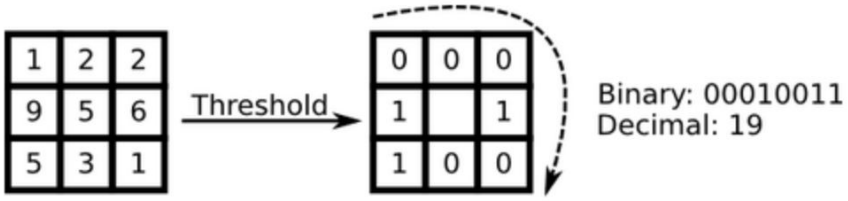

Figure 1. Example of LBP

The robustness of each method discussed will be assessed with a Mean Squared Error computation of each facial point, based on the method used to calculate it.

In this project, each member of the team will work together to perform the required analysis. Everyone will be responsible for looking over the dataset, writing the necessary code, and submitting the final report.

Expected Results

When all is said and done, we expect that we will be able to take any facial image of similar size to the training data and detect the facial features. Through analysis of the training data, and the clustering of features formed from it, we should be able to approximately determine the fifteen key facial points on a particular face.