Project Report

Face Recognition and PCA

Big Data Analytics

(CS 696)

Fall 2018

By

Reetu Hooda

A25244657

Table of Contents

1. Introduction

For humans, understanding and identifying emotions can be extremely interesting and useful, as genuine emotions are at most only partially controllable and often display their presence through facial expressions of the person experiencing them. A person’s emotions can sometimes be very distinct and obvious and at other times may very transient and difficult to notice; however, as long as their cues are visually present, it ostensibly possible for a computer to perform image processing and classification of that expression. There are many applications where the automated ability to process and detect emotion of a person can have functional benefits.

Interpersonal interaction is oftentimes intricate and nuanced, and its success is often predicated upon a variety of factors. These factors range widely and can include the context, mood, and timing of the interaction, as well as the expectations of the participants. For one to be a successful participant, one must perceive a counterpart’s disposition as the interaction progresses and adjust accordingly. Fortunately for humans this ability is largely innate, with varying levels of proficiency.

Humans can quickly and even subconsciously assess a multitude of indicators such as word choices, voice inflections, and body language to discern the sentiments of others. This analytical ability likely stems from the fact that humans share a universal set of fundamental emotions. Significantly, these emotions are exhibited through facial expressions that are consistently correspondent. This means that regardless of language and cultural barriers, there will always be a set of fundamental facial expressions that people assess and communicate with. After extensive research, it is now generally agreed that humans share seven facial expressions that reflect the experiencing of fundamental emotions. These fundamental emotions are:

1. Anger
2. Contempt
3. Disgust
4. Fear
5. Happiness
6. Sadness
7. Surprise

The universality of these expressions means that facial emotion recognition is a task that can also be accomplished by computers. Furthermore, like many other important tasks, computers can provide advantages over humans in analysis and problem-solving. Computers that can recognize facial expressions can find application where efficiency and automation can be useful, including in entertainment, social media, content analysis, criminal justice, and healthcare. For example, content providers can determine the reactions of a consumer and adjust their future offerings accordingly or health tracking apps that would monitor emotional stability and fluctuation of a user.

It is important for a detection approach, whether performed by a human or a computer, to have a taxonomic reference for identifying the seven target emotions. A popular facial coding system used both by noteworthy psychologists and computer scientists such as Ekman and the Cohn-Kanade group, respectively, is the Facial Action Coding System (FACS). The system uses Action Units that describe movements of certain facial muscles and muscle groups to classify emotions. Action Units detail facial movement specifics such as the inner or the outer brow raising, or nostrils dilating, or the lips pulling or puckering, as well as optional intensity information for those movements. As FACS indicates discrete and discernible facial movements and manipulations in accordance to the emotions of interest, digital image processing and analysis of visual facial features can allow for successful facial expression predictors to be trained.

2. Implementation

The workflow of the algorithm is summed up by the flowchart shown in Figure 1.

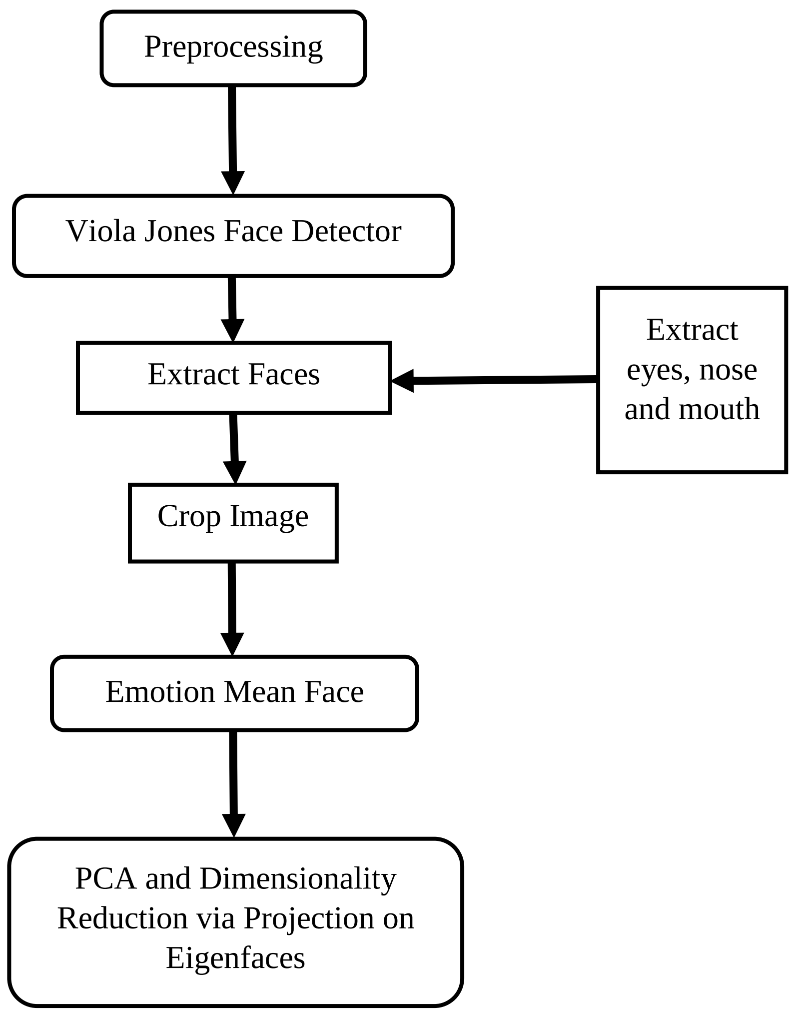

Figure 1: Flowchart of the algorithm.

It starts with preprocessing of the image followed by the overall face extraction from the image using a Viola-Jones cascade object face detector. Then faces are extracted and cropped out. We calculate mean image using the stacked cropped images to perform PCA and dimensionality reduction via projection on eigenfaces. Each of these steps are elaborated below:

Viola-Jones Face Detector: The Viola-Jones detection framework seeks to identify faces or features of a face (or other objects) by using simple features known as Haar-like features. The process entails passing feature boxes over an image and computing the difference of summed pixel values between adjacent regions. The difference is then compared with a threshold which indicates whether an object is considered to be detected or not. This requires thresholds that have been trained in advance for different feature boxes and features. Specific feature boxes for facial features are used, with expectation that most faces and the features within it will meet general conditions.

Essentially, in a feature-region of interest on the face it will generally hold that some areas will be lighter or darker than surrounding area. For example, it is likely that the nose is more illuminated than sides of the face directly adjacent, or brighter than the upper lip and nose bridge area. Then if an appropriate Haar-like feature, such as those shown in Figure 2, is used and the difference in pixel sum for the nose and the adjacent regions surpasses the threshold, a nose is identified. It is to be noted that Haar-like features are very simple and are therefore weak classifiers, requiring multiple passes.

However, the Haar-like feature approach is extremely fast, as it can compute the integral image of the image in question in a single pass and create a summed area table. Then, the summed values of the pixels in any rectangle in the original image can be determined using a total of just four values. This allows for the multiple passes of different features to be done quickly. For the face detection, a variety of features will be passed to detect certain parts of a face, if it were there. If enough thresholds are met, the face is detected.

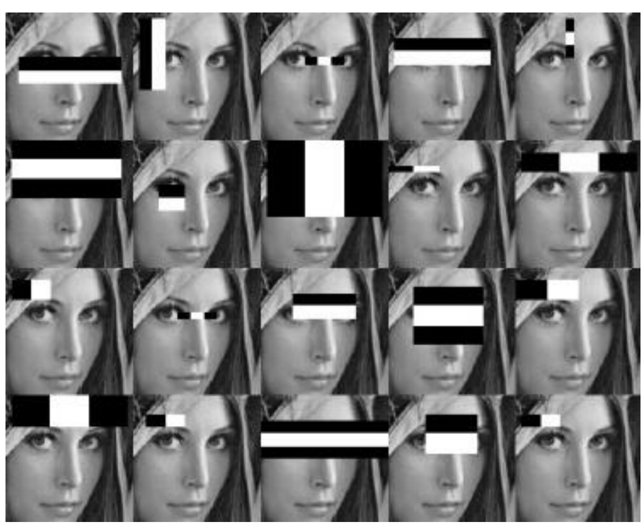

Figure 2: Sample Haar-like features for detecting face features.

Extract Faces and Crop Image: Once the faces are detected, they are extracted and resized to a predetermined dimensional standard. We will resize the extracted faces to 100x100 pixels. This will reduce computational demand in performing the further analysis.

Emotion Mean Face: The mean image for all faces will be calculated. The entire training set is comprised of faces from the Extended Cohn-Kanade dataset and comprises faces that express the basic emotions. The mean image is then subtracted from all images in the training set.

Principal Component Analysis: Using the mean-subtracted training set the scatter matrix **S** is formed. The intention is to determine a change in basis that will allow us to express our face data in a more optimized dimensionality. Doing so will allow the retention of most of the data as a linear combination of the much smaller dimension set. PCA accomplishes this by seeking to maximize the variance of the original data in the new basis. We perform PCA on the using the Sirovich and Kirby method, where the eigenvalues and eigenvectors of the matrix **S**H**S** are first computed to avoid computational difficulties. The eigenvectors of the scatter matrix, defined as **SS**H, can then be recovered by multiplying the eigenvector matrix by **S.** Retaining the top eigenvectors, also known in this context as eigenfaces, allows us to project our training data onto the top eigenfaces, in this case the 100 associated with the top eigenvalues, in order to reduce dimensionality while successfully retaining most of the information.

3. Results

Figure 3 shows one of the sample images from the dataset.

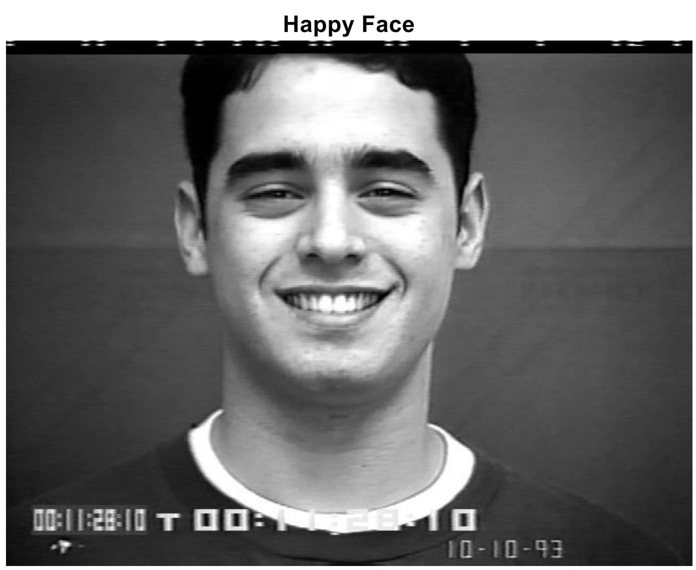

Figure 3: Sample Image

The face is extracted from the sample image using Viola-Jones cascade object detector as shown in Figure 4. Extracted face is cropped shown in Figure 5.

Happy Detected face

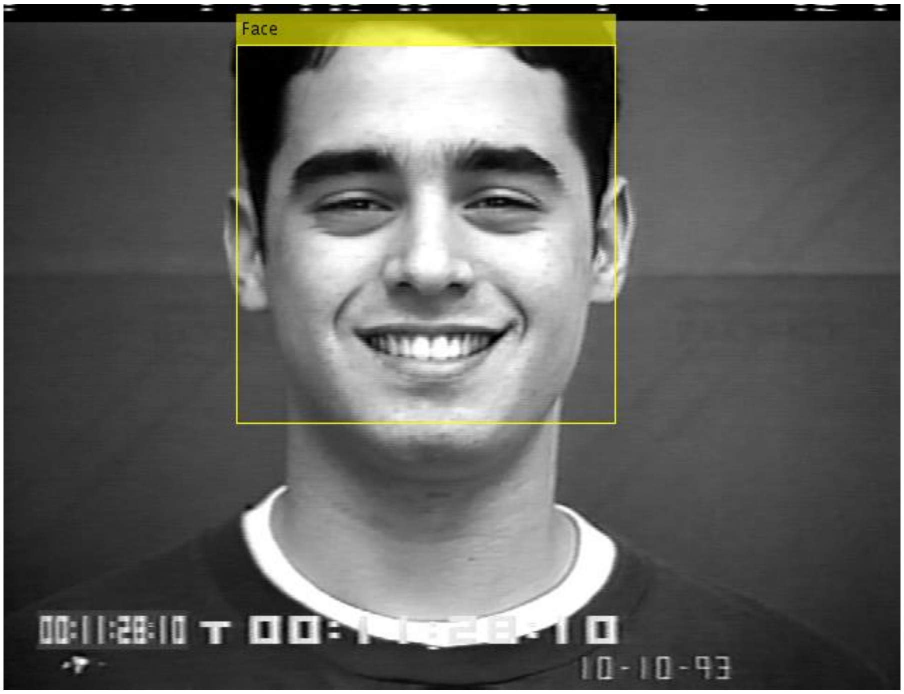

Figure 4: Face Detection

Cropped Face

Figure 5: Cropped Image (100X100)

Figure 6 shows the emotion mean face obtained by calculating the mean of stacked images for a specific emotion. Finally using PCA we find out the eigenfaces shown in Figure 7.

Figure 6: Emotion Mean Face

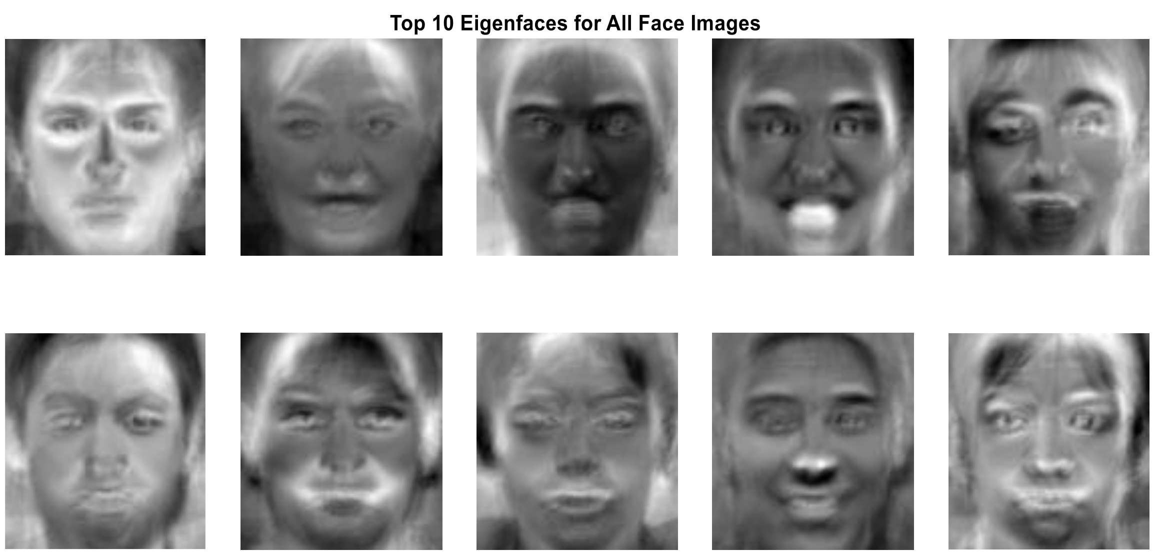

Figure 7: Eigenfaces using PCA

4. Conclusion

An image processing methodology for face detection and PCA on the detected images is implemented in this project. The image processing involved steps to clean, color balance, and appropriately segmentation of the data. Certain features were detected, such as the eyes, lips, and teeth. Those features are isolated with segmentation, and inter-feature distances, angles, and curvatures are measured. The completed training implementation uses Viola-Jones's Haar-like feature cascade detector to detect faces as well as eyes and mouths. Detected faces are cropped, resized, and mean subtracted, then PCA is performed to obtain the eigenfaces.

The next project should include the Fisher linear discriminant analysis (LDA) in a reduced dimensionality. For each emotion that we wish to train a predictor for, we will perform Fisher LDA, in which the goal is to optimize the objective function that minimizes within class variance and maximizes between class variance to gain clear class separation between the classof interest and the other classes.

References

[1] Lucey, P., Cohn, J.F., Kanade, T., Saragih, J. Ambadar, Z. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. IEEE Computer Society Conference CVPRW (2010).

[2] T. Mita, T. Kaneko, O. Hori, Joint Haar-like Features for Face Detection, "Proceedings of the Tenth IEEE International Conference on Computer Vision", 1550- 5499/05 ©2005 IEEE

[3] Computer Vision", 2004. [9] M. A. Turk and A.P. Pentland, Face recognition using eigenfaces, "Proceedings of the IEEE", 586-591, 1991.