Project Report

Emotion Detection through Facial Feature Recognition

Big Data Analytics

(CS 696)

Fall 2018

By

Reetu Hooda

A25244657

Table of Contents

1. Introduction

(a) Summary of Project 1

An image processing methodology for face detection and PCA on the detected images was implemented in project 1. 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 were measured. The completed training implementation used Viola-Jones's Haar-like feature cascade detector to detect faces as well as eyes and mouths. Detected faces were cropped, resized, and mean subtracted, then PCA is performed to obtain the eigenfaces.

(b) Summary of Project 2

Second project included the Fisher linear discriminant analysis (LDA) in a reduced dimensionality. For each emotion that we wish to train a predictor for, we performed Fisher LDA, in which the goal was to optimize the objective function that minimizes within class variance and maximizes between class variance to gain clear class separation between the class of interest and the other classes.

(c) Project 3

As mentioned in the future work of project 2, I wished to develop another classifier in addition to our Fisherface based classifier since, as we find out experimentally, the Fisherface approach is limited in success by itself. I leveraged onto the fact that most expression information is encoded within the inner facial features, specifically the regions around the eyes, nose, and mouth.

2. Implementation and Results

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

As is detailed in FACS, the inner facial features will move in certain distinct combinations with the exhibition of each emotion, as is described by Action Units. Visually, these movements and manipulations should be evidenced in changes of gradients in the areas in the inner facial features. In particular, the brows and mouth, and how they visually warp, are very important in the detection of emotions. We will utilize this information to train a classifier which can predict emotions based on the information encoded in the gradients.

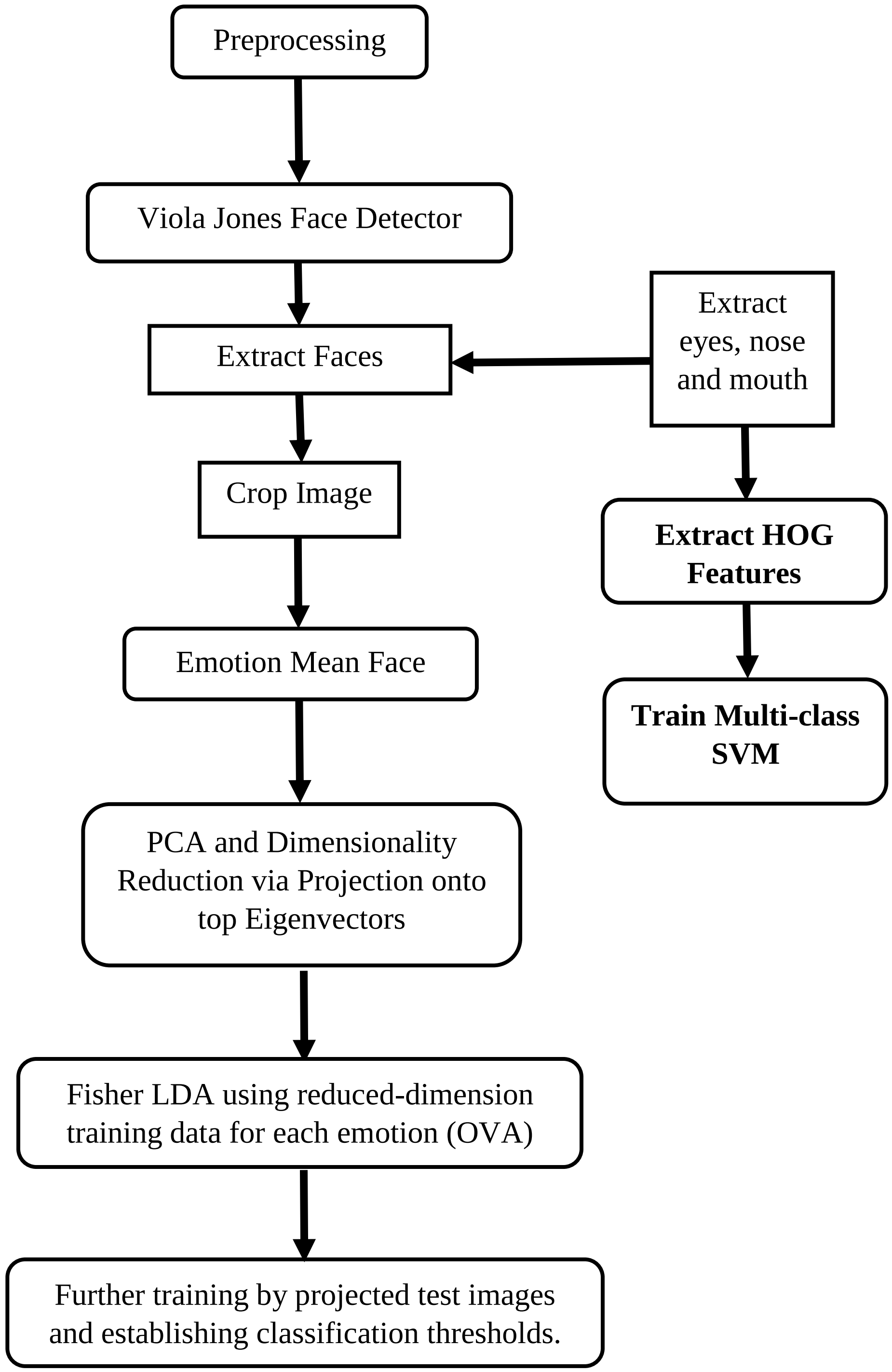

Figure 1: Flowchart of the algorithm.

To begin, we must first extract the eye and mouth regions. We first try to detect these features separately using Haar-like features again. This approach is mostly successful. However, when it is not, perhaps due to illumination issues that affect the Haar-like feature calculations and the thresholding, we need another approach.

Here we propose the use of Harris corner detection to detect features such as the eyes in a face image. The Harris corner detection method seeks to find points in an image that are corners by the definition that moving in any direction from that point should provide a gradient change. The approach is to use a sliding window to search for the corner points by examining gradient changes when sliding across that area.

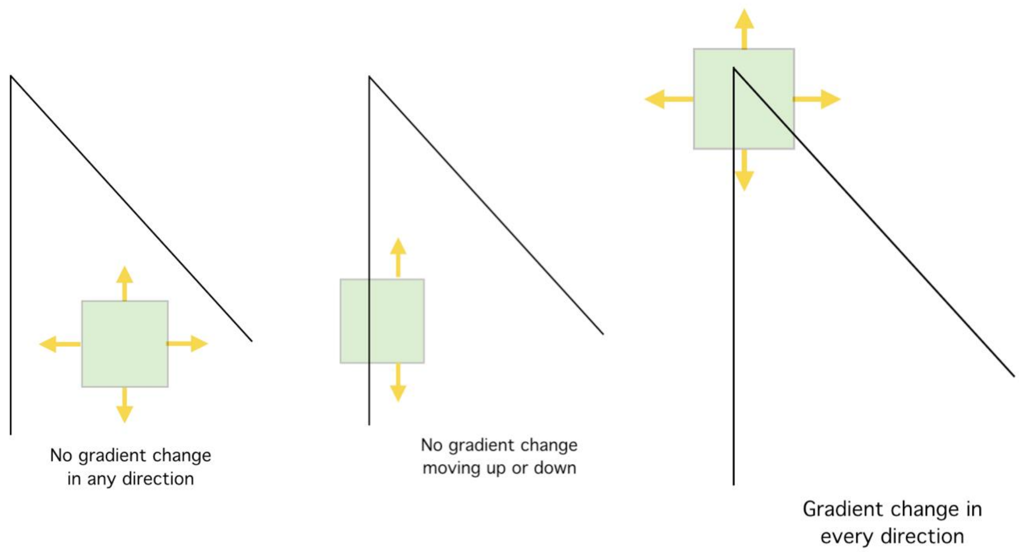

Figure 2: Harris corner detection method, where a corner is detected if moving in every direction from that point results in a large gradient change.

We use the fact that the eyes in a face image will be very non- uniform relative to the rest of the face. There white portion of the human's eye is surrounded by skin that is darker, and the pupil and iris in the center of the eye is almost always darker as well. When viewing a face image with varying pixel intensities, some of the strongest corners are in the eye region. We use this fact to find eyes in a face when the Haar-like feature approach fails.

Figure 3 gives an idea of the Harris corner extraction approach. We find the Harris corners on a cropped face image, then keep a number of the strongest corners. We then partition the face into vertical intervals and tally the number of Harris corners that fall in that vertical interval. The interval with the most Harris corners detected "wins the vote" and the eyes are determined to fall in that interval. From that information, the eyes are then extracted.

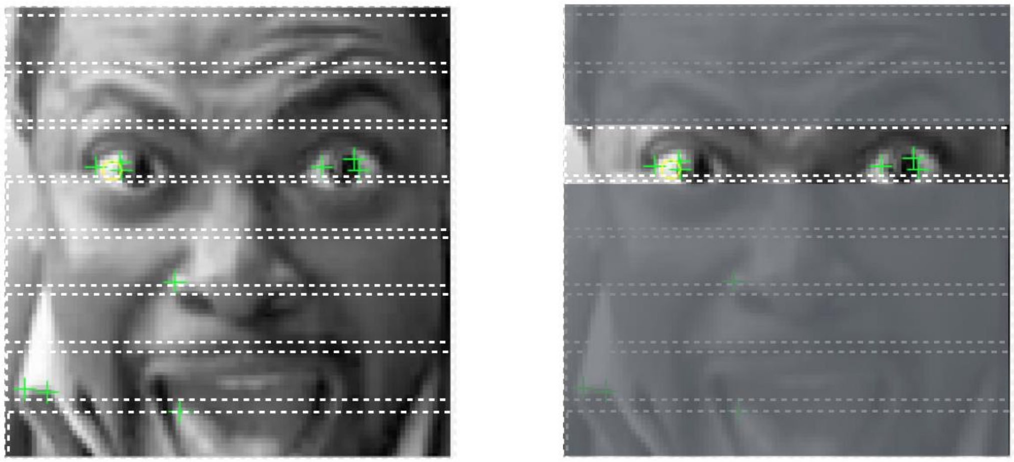

Figure 3: Harris corner approach for feature extraction, where the strongest cornerpoints are shown as green crosses. The corner locations are tallied in vertical intervals and the interval in which the eyes reside is determined.

Following extraction of the eyes and the mouth regions, HOG features are calculated and extracted. To determine the HOG features, an image is separated into evenly-sized and spaced grids. Within each grid, the orientation of the gradient for each pixel at *(x,y)* is calculated as:

$$\theta_{x,y} = \tan^{-1} \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}$$

where L is the intensity function describing the image. These orientations of gradients are then binned into a histogram for each grid, and every grid within the image is concatenated resulting in a HOG description vector. Figure 4 shows examples of calculated HOG features that are plotted on top of the image regions that they correspond to.

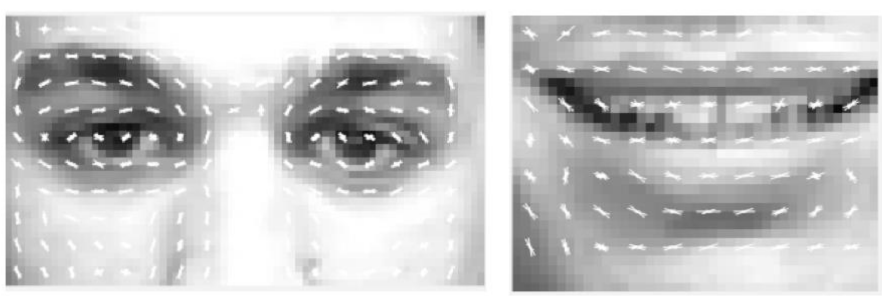

Figure 4: Plotted visualizations of HOG features on extracted eye and mouth regions.

It should be expected then that facial expressions that have different muscular manipulations should result in varying HOG features. It should be noted that the extracted and resized eye and mouth regions must be consistent in dimension from image to image so we can extract the same number of HOG features, which is required for our further classifier training.

We concatenate the extracted eye HOG vector with the mouth HOG vector for each training image, and assign a corresponding label. This, like the Fisher LDA process, requires us to know the class that each test image belongs to. Upon completing HOG extraction for each image, we then train a mulit-class support vector machine (SVM) using the concatenated HOG vector.

The testing results on test images from Cohn-Kanade using MATLAB are given in Table 1.

Algorithm	Accuracy
Fisherface only (Angry, Fear, Sad, Surprise, Happy)	90%
Fisherface only (All 7 emotions)	56%
HOG only (All 7 emotions)	81%
Fisherface + HOG (All 7 emotions)	81%

Table 1: Testing results for classifier.

3. Discussion

During training, eye and mouths are detected using Haar-like features, or using a Harris corner based approach is Haar-like features fail. The detected eye and mouth regions are then extracted and resized. HOG features are extracted from each region, and a SVM is trained using a combined eye-mouth HOG vector and training labels.

The primary reason we use this dual-classifier approach is improving speed with maintaining accuracy. The dual-classifier approach works well when the Fisherface cannot effectively determine a prediction. This happens in two cases. First is if a test image is not one of the "easy-to-distinguish" emotions, and second is if the Fisherface classifier cannot decide between two or more predicted emotions.

The overall testing approach is to pass a test image through each of the five "easy-to-distinguish" Fisherface classifiers. If only one classifier makes a positive prediction, then that test image is assigned that Fisherface's emotion as the prediction.

If no classifier offers a positive prediction, or more than one classifier offers a positive prediction, then the test image moves to phase two of the classification processes. The test image first undergoes Haar-like feature detection for the eye region and mouth region. The detailed Harris corner method is used as a backup if Haar detection fails. Then the HOG features are extracted for both regions, concatenated, then passed to the trained SVM for a final prediction.

When using the HOG and SVM classifier only, the accuracy for detection is 81%, much better than a Fisherface only approach. When using the dual-classifier method, the accuracy is the same as HOG-only at 81%, but the testing process is 20% faster. This is because not all images must undergo eye and mouth detection, extraction, then undergo HOG feature extraction, but only those test images that are not given a prediction by the much faster Fisherface classifier.

4. Conclusion

An image processing and classification method has been implemented in which face images are used to train a dual- classifier predictor that predicts the seven basic human emotions given a test image.

The predictor is relatively successful at predicting test data from the same dataset used to train the classifiers. However, the predictor is consistently poor at detecting the expression associated with contempt. This is likely due to a combination of lacking training and test images that clearly exhibit contempt, poor pre-training labeling of data, and the intrinsic difficulty at identifying contempt. The classifier is also not successful at predicting emotions for test data that have expressions that do not clearly belong exclusively to one of the seven basic expressions, as it has not been trained for other expressions.

Future work should entail improving the robustness of the classifiers by adding more training images from different datasets, investigating more accurate detection methods that still maintain computational efficiency, and considering the classification of more nuanced and sophisticated expressions.

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.