Report Project 2

Name: Alhiet Orbegoso

Unity ID: 200322491

Class: ECE 763 Computer Vision

A. Introduction

The project objective is to classify an image as face or non-face. This task is executed using the AdaBoost algorithm. AdaBoost algorithm creates a strong classifier using a linear combination of weak classifiers where each one alone has an accuracy slightly above the random guess (0.5).

As weak classifiers $h(x)$ it has been used the face Haar features defined in Viola-Jones paper [1]. The weak classifier predict face if $h(x) = +1$ and non-face if $h(x) = -1$. Each weak classifier is defined as:

$$h(x) = \begin{cases} +1 & if\ f(x) \geq T \\ -1 & if\ f(x) < T \end{cases}$$

Where, $f(x)$ is the absolute value of the difference between pixels inside in white and black regions respectively:

$$f(x) = abs\left(\sum_{p \in White} p(x,y) - \sum_{p \in Black} p(x,y)\right)$$

For this project the value of T has been set to 5. The value of T depends on the quantity of pixels inside a Haar feature and its value range ([0,1] or [0,255]). In this project the range has been set to [0,1]. The value selected increase the probability to select Haar features which have similar size as face features.

At each iteration t, it will be evaluated each weak classifier prediction over all samples. The classifier with the lowest error prediction is selected. Then alpha α_t is calculated using the following formula:

$$\alpha_t = \frac{1}{2} \ln\left(\frac{1 - e_t(h_t)}{e_t(h_t)}\right)$$

The algorithm used it is detailed in table 1 of Viola-Jones paper [1].

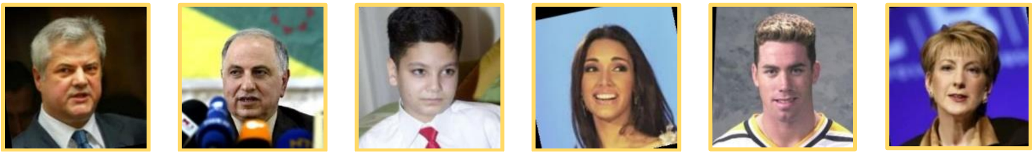
Finally, the strong classifier has the form:

$$H(x) = \sum_{t=1}^{T} sgn\big(\alpha_t . h_t(x)\big)$$

B. Dataset

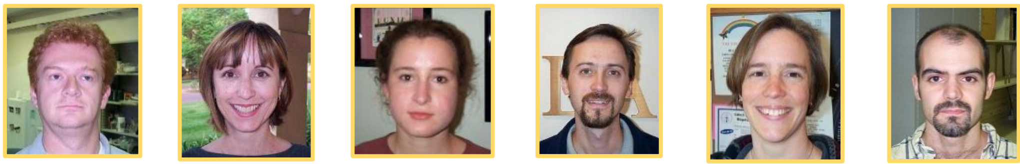
The dataset is ensembled from 2 sources, dataset **Labeled Faces in the Wild** (http://vis-www.cs.umass.edu/lfw/) from University of Massachusetts Amherst and dataset **Caltech 101** (http://www.vision.caltech.edu/Image_Datasets/Caltech101) from California Institute of Technology.

- **Labeled Faces in the Wild:** This dataset provides aligned face images from uncontrolled environment, challenging the classification models. From this dataset has been extracted 1100 face images.



Face samples from dataset

- **Caltech 101:** This dataset provides aligned faces from a more controlled environment. Most faces are front view, same illumination and typically represent the 70% of pixels. From this dataset has been extracted 435 face images.



Face samples from dataset

In addition, this dataset provides the non-face images: random objects photos and drawings. From this dataset has been extracted 1535 non-face images.

Non-faces samples from dataset

C. Experimental setup

All images have a preprocessing step. First are converted to gray images, then a histogram equalization processing is performed and finally the size is reduced to 24x24.

For this experiment has been selected randomly 1200 faces images for training and 300 for testing. In the same way it has been selected 1200 non-faces images for training and 300 non-face images for testing.

The code has been developed in Python 3.7.4; libraries installed are listed below:

- Numpy 1.17.1
- Matplotlib 3.1.2
- Scikit-image 0.16.2: Used for Haar features generation and integral image.

Given that haar features generated are above the 160000. A preprocessing step has been added to the code. The objective is to reduce the number of features by selecting the ones with pixels below a threshold 1 and above a threshold 2.

Haar features with few pixels commonly have high error rates and are discarded by algorithm so it would be a waste of resources to process them.

Haar features with many pixels will are less likely to capture the details of the face so they are also discarded.

For this project it has been imposed a low threshold of 40 pixels and a high threshold of 520 pixels. The number of features is reduced from 161864 to 96666 (40% reduction).

The algorithm is executed with 20 iterations. The final strong classifier has 20 weak classifiers.

The following table presents a summary of parameters used for the experiment:

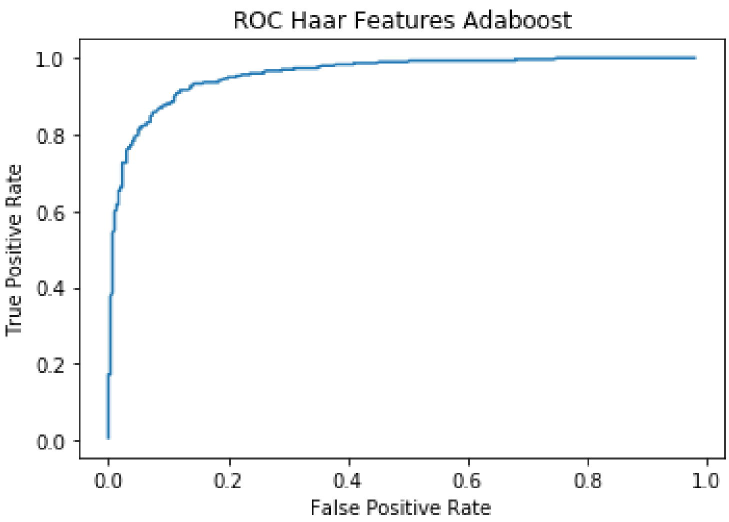
Parameter	Value
Image Size	24x24
Histogram Equalization	YES
Train Size (Face)	1200
Train Size (non-Face)	300
Test Size (Face)	1200
Test Size (non-face)	300
Weak Classifier Threshold	5
Low Limit Haar Pixels	40
High Limit Haar Pixels	540
Iterations	20

D. Results

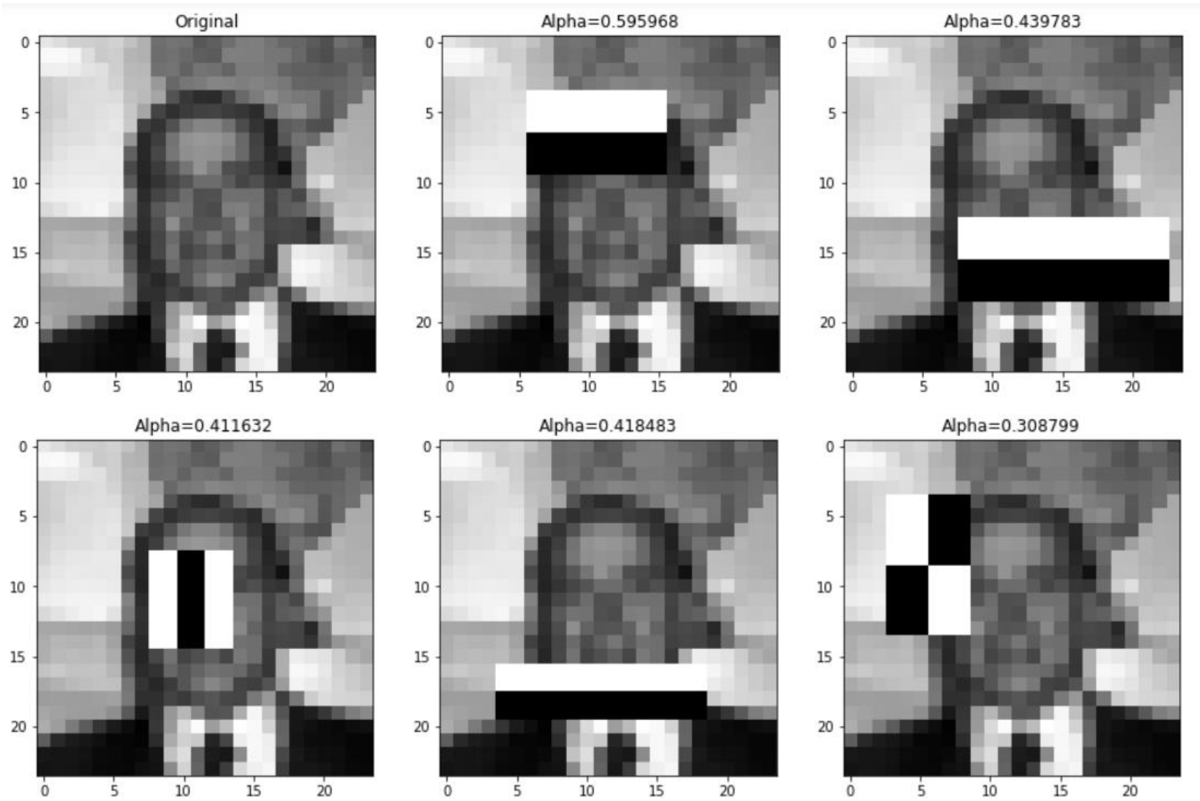
Following is presented the true positive ratio, true negative ratio, accuracy and ROC curve over test samples.

True Positive Rate	0.88
True Negative Rate	0.9
Accuracy	0.89

Performance Rates



Here is also presented some Haar features calculated from algorithm:



E. Results (Reference Dataset)

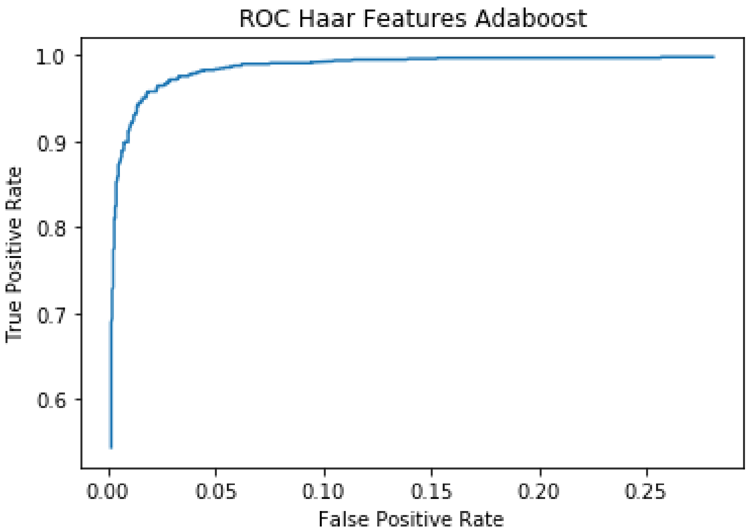
Parameters summary

Parameter	Value
Image Size	16x16
Histogram Equalization	YES
Train Size (Face)	4100
Train Size (non-Face)	8000
Test Size (Face)	1000
Test Size (non-face)	2000
Weak Classifier Threshold	5
Low Limit Haar Pixels	8
High Limit Haar Pixels	220
Iterations	20

Following is presented the true positive ratio, true negative ratio, accuracy and ROC curve over test samples.

True Positive Rate	0.945
True Negative Rate	0.986
Accuracy	0.972

Performance Rates



F. References

[1] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Kauai, HI, USA, 2001, pp. I-I.