Viola Jones Face Detection Project with AdaBoost

Jincheng Li

ECEN 649 HW5

Nov 23, 2020

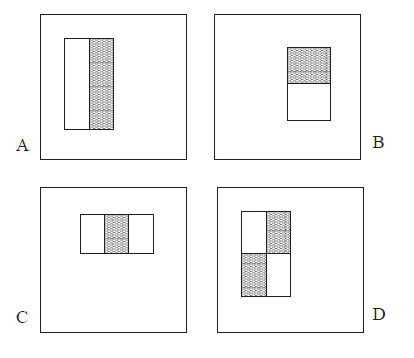
**Methodology**

The purpose of this project is to implement the Viola-Jones face detection algorithm with AdaBoost. The training cases provided are pictures of faces and non faces images of 19 \* 19 resolution gray scale images. In the training case, 499 faces and 2000 non faces are provided while in the testing case, 472 faces and 19572 non faces are included.

The main procedures are divided into 3 parts: extracting Harr features; implementing ERM decision stumps on the training set; implementing AdaBoost algorithm.

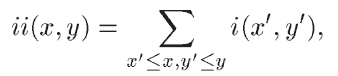
**Extracting Harr Features**

In the Viola Jones paper, the Harr features is defined as a specific rectangular area selected from the image. The sum of all the pixel values from the white part minus the sum of all the pixel values in the black part is calculated as the Harr feature value. In Fig. 1, Harr features 1 – 4 are defined as the construction of A, B, C, D.

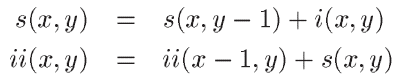


**Fig. 1 – Harr feature construction**

Since for each image, various locations, sizes, and types of Harr features needs to be selected to calculate the Harr feature value, it would be a demanding and inefficient task to adds up all the pixel values repeatedly. Therefore, the paper introduced the idea of integral image, which is defined as Fig. 2 shown below. The integral image at location (x, y) contains the sum of all the pixel above and to the left of (x, y), inclusively.

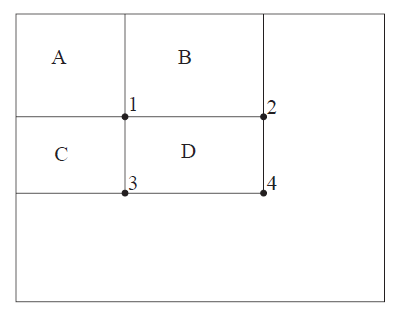


**Fig. 2 – Integral image value at (x, y) equation**



**Fig. 3 – s(x, y) is the cumulative row sum, ii(x, y) is the integral image and i(x, y) is the original image.**

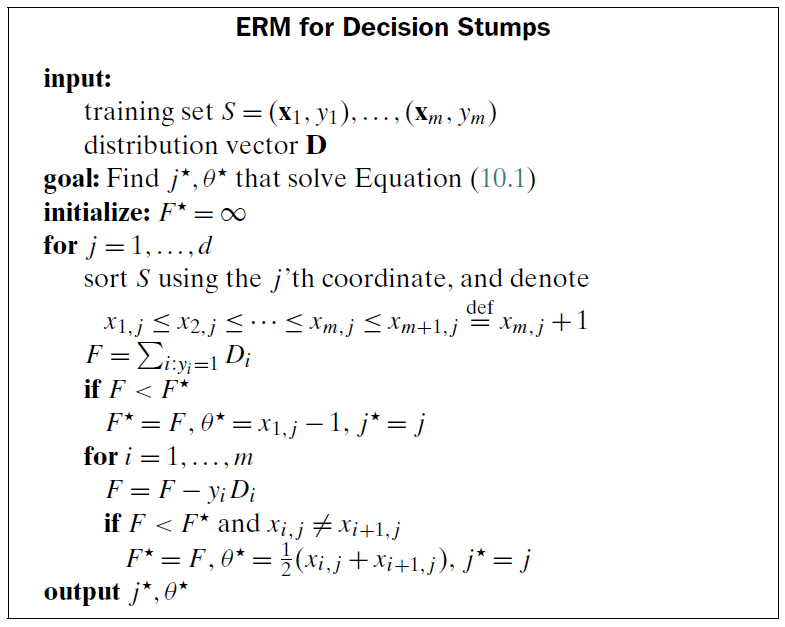
With above equations, it would be very easy to calculate the sum of all pixel value at a specific area. For example, in the Fig. 4, the sum of pixel in D can be calculated as the integral image at point 4 + integral image at point 1 – integral image at point 2 – integral image at point 3.



**Fig. 3 - The sum of the pixels within rectangle D can be computed with four array references. The value of the integral image at location 1 is the sum of the pixels in rectangle A. The value at location 2 is A + B, at location 3 is A + C, and at location 4 is A + B + C + D. The sum within D can be computed as 4 + 1 - (2 + 3)**

**ERM Decision Stumps**

After extracting the Harr feature values, I assigned it into the training set as x1 to xm. The faces trained has label of 1 as y while the non faces are assigned -1. The decision stumps are executed with the following algorithm. What it does is it selects the one feature that can minimize the empirical risks. In another word, it chooses one Harr feature each round as j\* and calculated threshold θ\*. Using the threshold and Harr feature selection we can obtain the weak learner h, eventually be used in the AdaBoost each round.



**Fig. 4 – ERM Decision Stumps pseudocode**

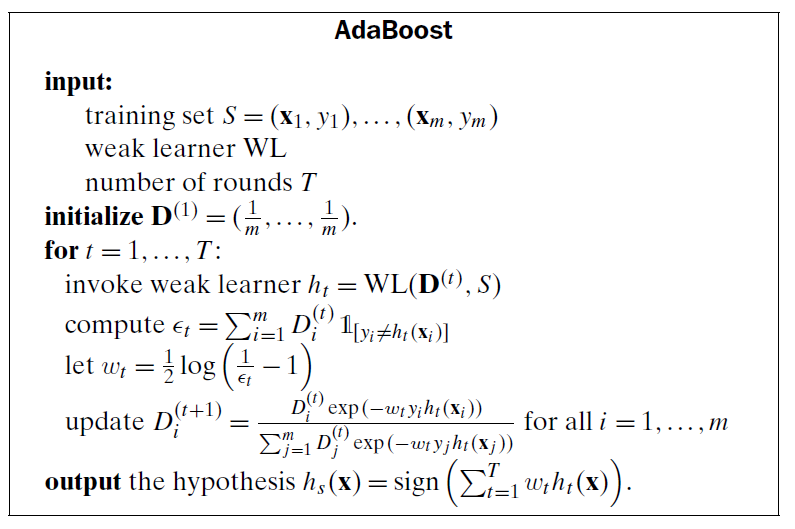


**Fig. 5 – Weak Learner with feature and threshold**

**AdaBoost**

First, all the training samples are initialized a uniform distribution of 1/D, where D is the total number of samples. The AdaBoost algorithm is run a total of 10 rounds, where in each round the ERM decision stumps are implemented to obtain the Harr feature selected and threshold. With this feature and threshold, the weak learner generates labels of prediction. The indices of output predictions where they don’t match the labels are collected, and the corresponding distribution value are added up to compute the epsilon factor.

Next, the round weight is calculated as shown in Fig. 6, where if the number of non matches are large, the round weight will decrease to compensate for it. In each round it will put more weight on where it does not predict well. Finally, the overall hypothesis is a combination of all the round weight multiply by the round-based weak learner prediction.



**Fig. 6 – AdaBoost Implementation**

**Result and Discussion**

1. The top 10 features selected by AdaBoost

In each AdaBoost round, the ERM decision stumps calculated j\* and θ\*, where j\* represent the column, corresponds to what type of Harr feature, the top left and the bottom right point of the rectangle of this feature. Since selecting Harr feature is reparative algorithm that iterates over length and width of certain sizes, the Harr feature map is the same for all images training or testing. I obtained the Harr feature information from the j\* to my Harr feature map. The top 10 features in each round is shown as following. It has been plotted on top of the face images for the purpose of visualization.



**Fig. 7 – Top 10 features selected by AdaBoost algorithm**

**Fig. 8 - Feature 1**

****

**Fig. 9 - Feature 2**

****

**Fig. 10 - Feature 3**

****

**Fig. 11 - Feature 4**

****

**Fig. 12 - Feature 5**

****

**Fig. 13 - Feature 6**

****

**Fig. 14 - Feature 7**

****

**Fig. 15 - Feature 8**

****

**Fig. 16 - Feature 9**

****

**Fig. 17 - Feature 10**

1. **The combined classifiers after running 1, 3, 5, and 10 boosting rounds.**

In each round the round weight, threshold, and polarization is listed in the following table. The combined classifier is a signed weighted sum of round weight multiply by predicted label using each round’s weak classifier shown in Fig. 9 below.



**Fig. 9 – Finally classifier as a combination of weak classifier in each AdaBoost round**

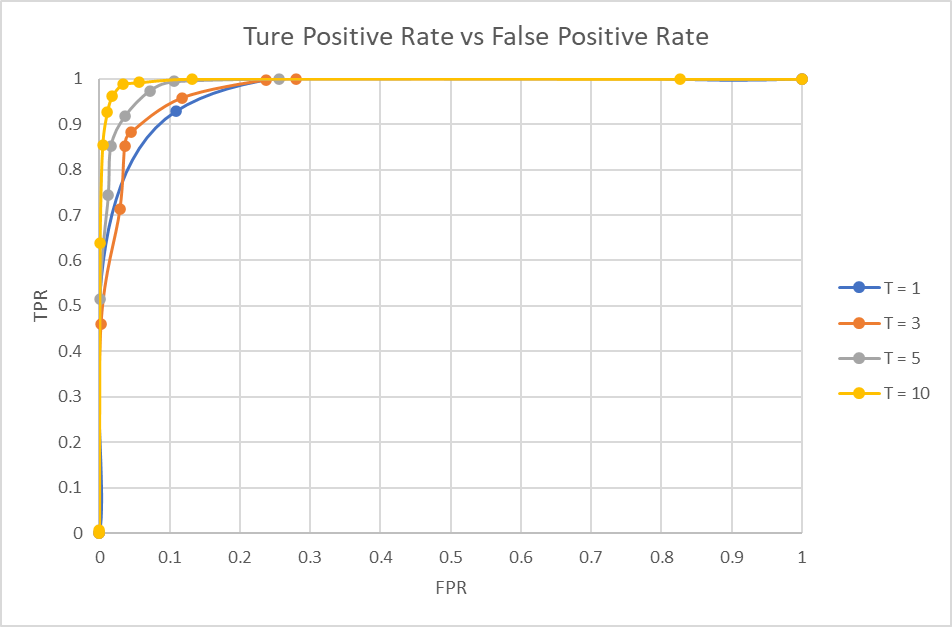
Fig. 10 below shows the tabulated round weight wt and threshold θt.



**Fig. 10 – Combined classifiers details of each round**

1. **The ROC curve of the combined classifiers after running 1, 3, 5, and 10 boosting rounds when applied to the test set.**

The ROC curve is obtained by adjusting the overall theta value. By default, it was set as 0, which is obtaining the sign of the overall classifier. I adjusted it manually to obtain different FPR (false positive rate) and FNR (false negative rate) and plot it as shown in Fig. 10.



**Fig. 11 – Tradeoff between True Positive Rate and False Positive Rate**

As we can see, the curve is shifted left, upwards as more AdaBoost round is performed. This is because as weaker classifier is combined, the overall prediction will have more systematic information in determining whether an image is a face or not. Therefore, the trade off between FPR and FNR is better as round increases.