STAT 231 Project II

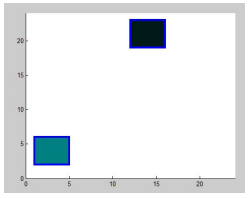
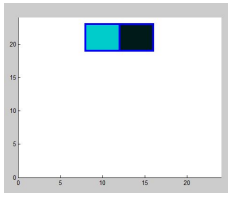
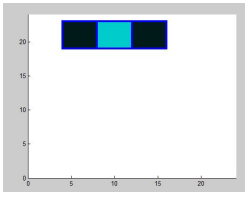
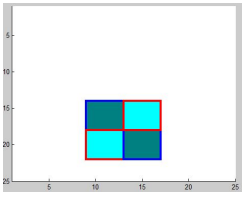
Detecting Faces in Images by Boosting Technique

Pei-Chun Su

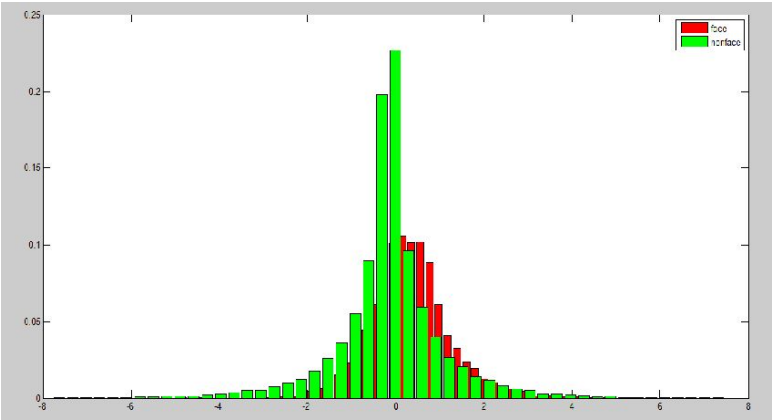
704517656

1. Construction of weak classifiers

I deisinged about 45000 weak classifiers as following

Type 1	Type 2	Type 3	Type 4
			

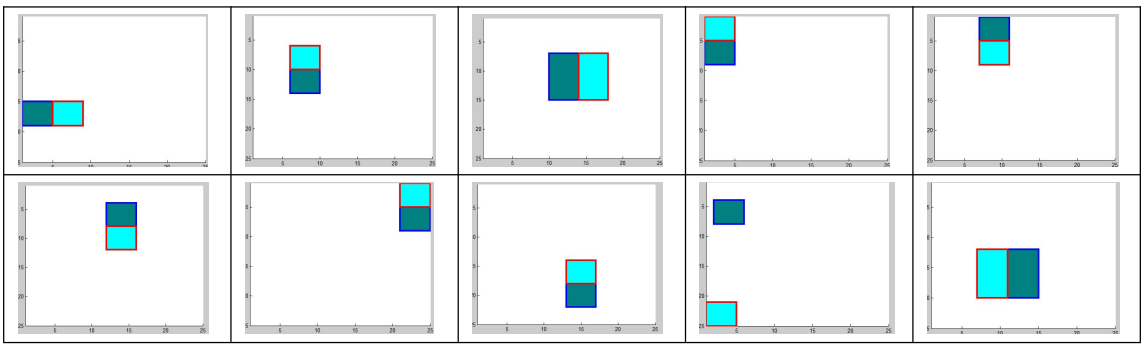
The first type is random two rectangles combination with one positive and one negative square.
The second type is two concatenated squares with different sign.
The third type is 3 square with two having same sign and the other one having differnt sign.
Each filter can change differnt sizes, shape of rectangles, and have transverse or vertical directions.
The hitogram of weak classifers are almost like following:



Each classifiers are weak. I decide the threshold of each classfier by the average of the 50th percentile filtered values of face and nonface groups.
Afterwards,the filtered image values are calculated by image integral techniques.

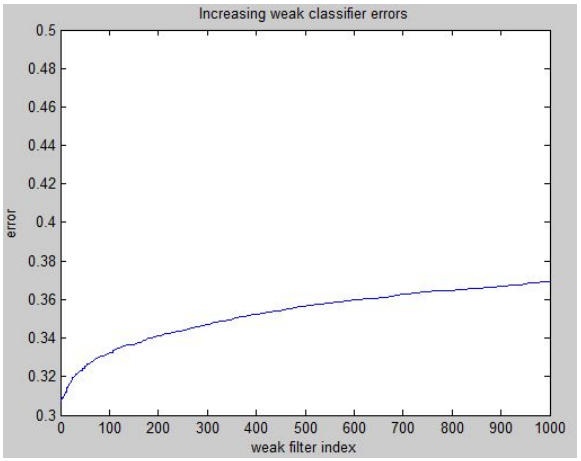
2. AdaBoosting

After implement the Adaboost algorithm to boost the weak classifiers in 1.
The best ten features are:

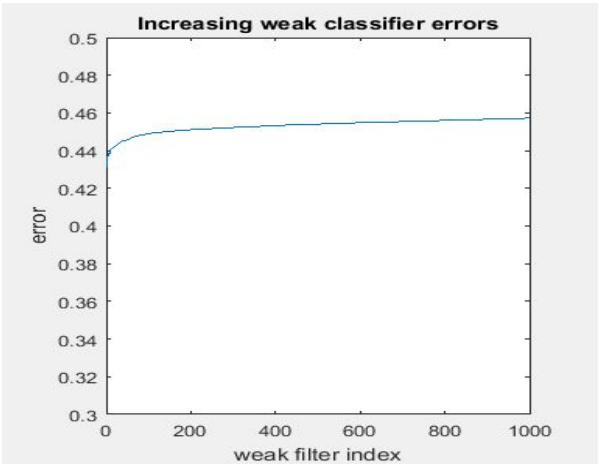


The following is the curve of errors, at steps T=0, 10, 50, 100 respectively, of top 1000 weak classifiers among the pool of weak classifiers in increasing order.

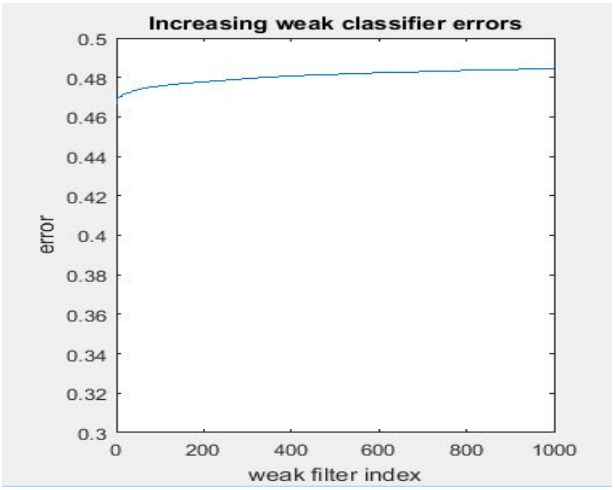
T = 0



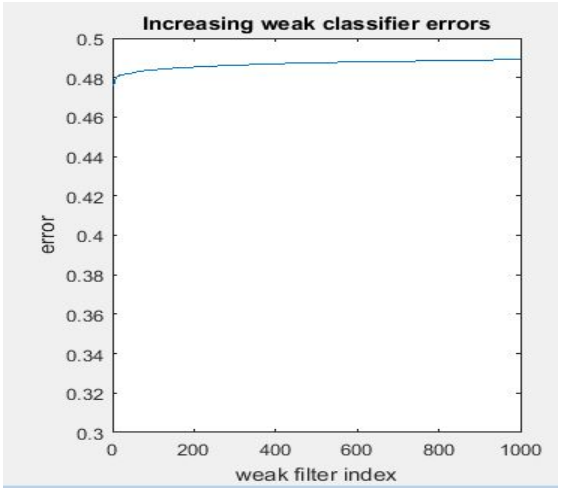
T=10



T-=50



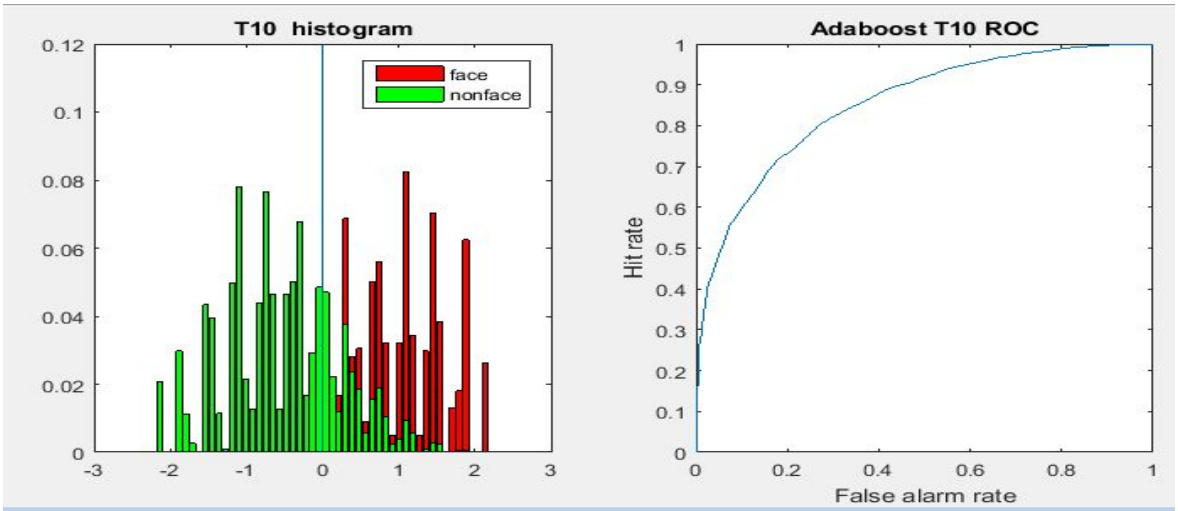
T=100



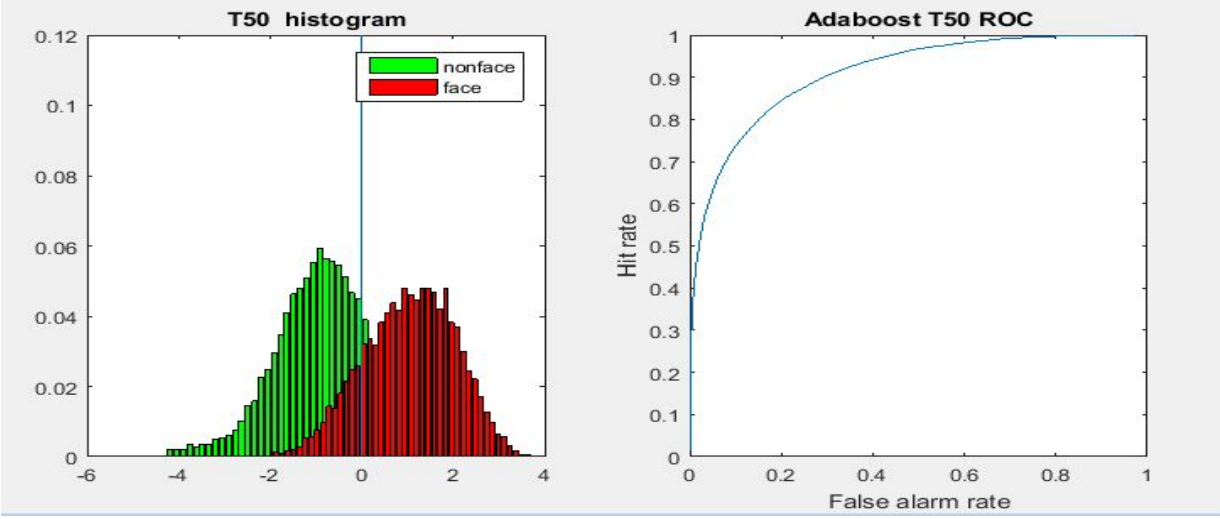
With T increase, the error of weak classifiers in pool arises and approach to 0.5, because the Adaboost algorithm put more weight on missclassified samples.

The Adaboost histograms of the positive and negative populations over the F(x) axis and the three corresponding ROC curves for T=10,50,100.

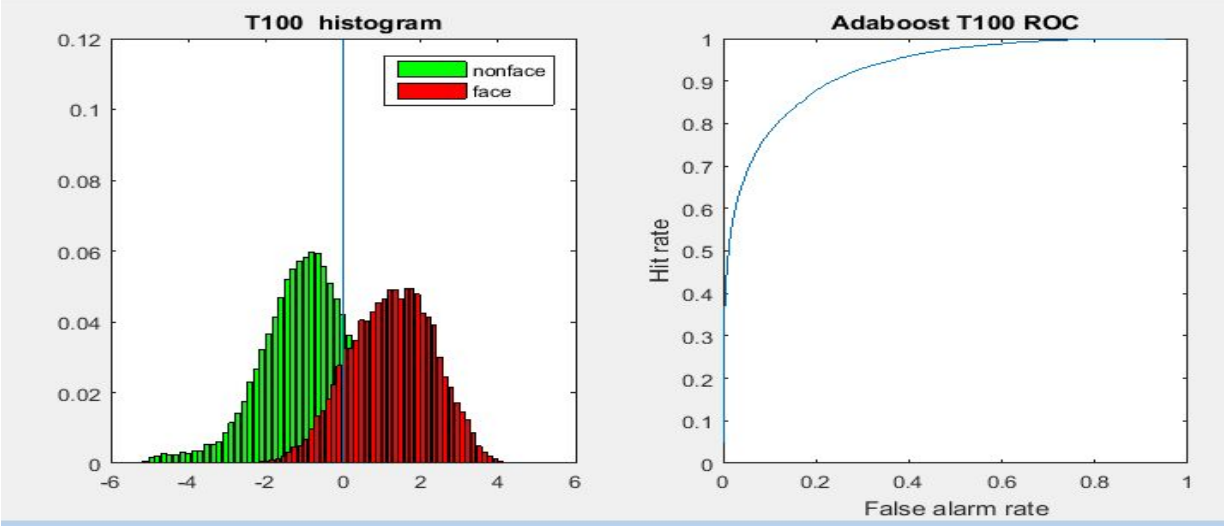
T = 10



T=50



T=100



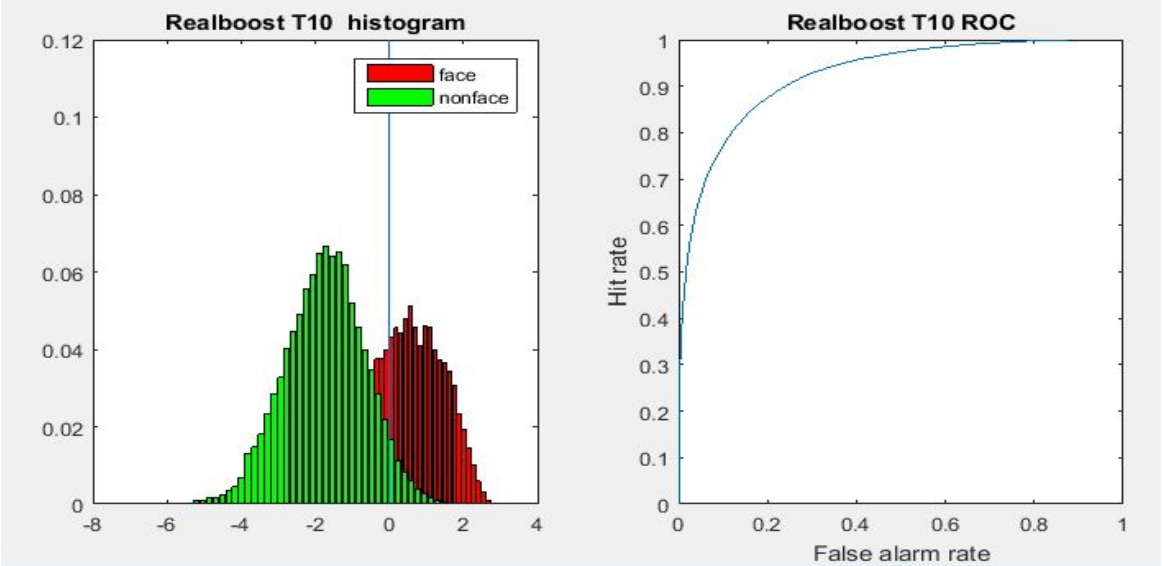
As T increase, the result final filter F(x) can better seperate positive and negative population. And the corner of the ROC curve move to left top more.

3. RealBoosting

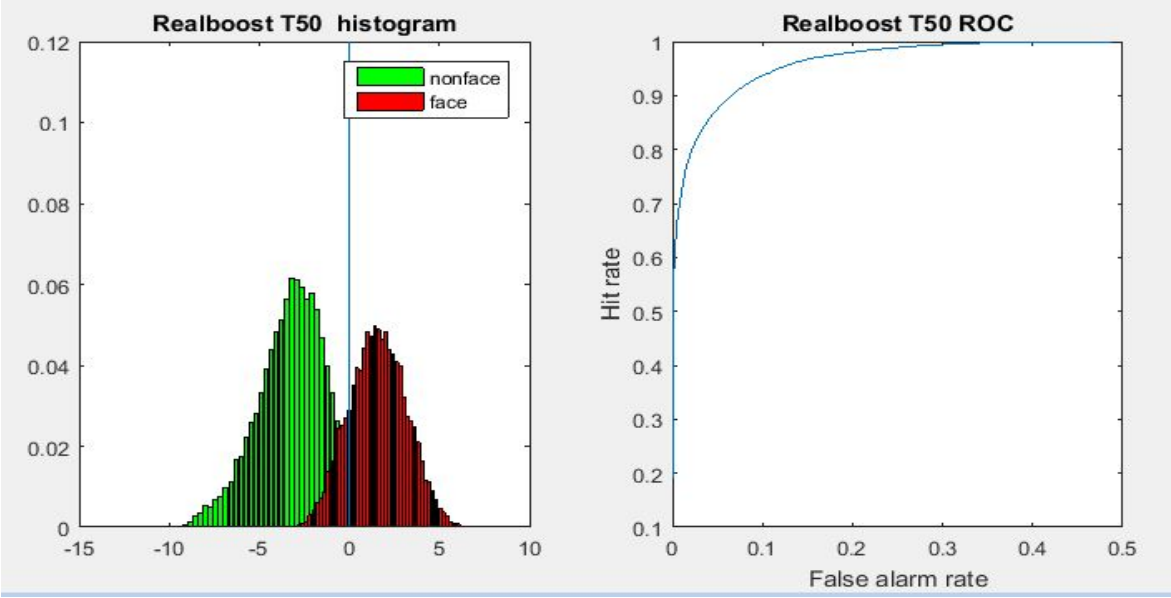
I use the T 10, 50, 100 features I got in step 2 of Adaboost to implement RealBoosting algorithm. I divide each filter to 40 bins, from the 5th percentile of positive population to 95th percentile of negative population.

The Realboost histograms of the positive and negative populations over the F(x) axis and the three corresponding ROC curves for T=10,50,100.

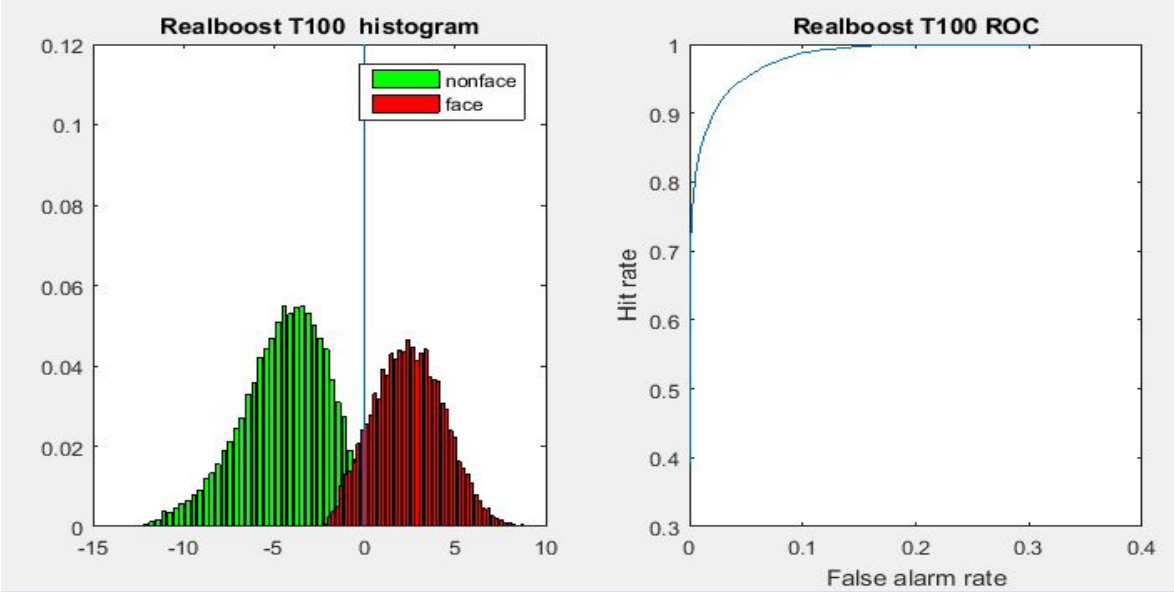
T=10



T=50



T=100

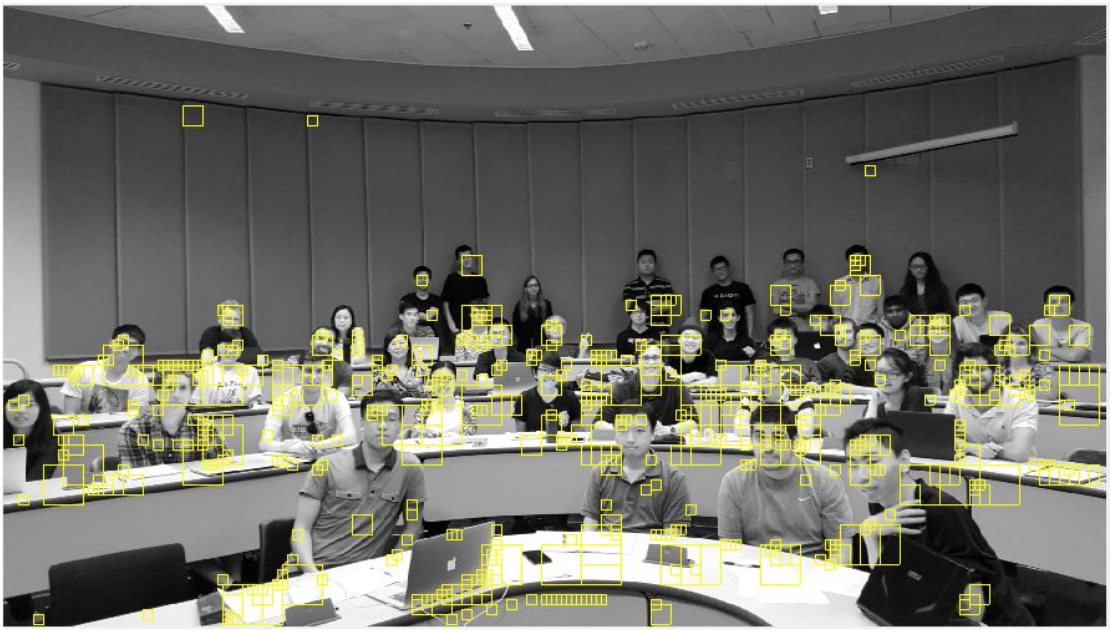


Compare to Adaboost, Realboost seperate the positive and negative populations better and have better ROC curve with the same T. This shows that Realboost algorithm converge to the error upper bound faster then Adaboost.

4. Test images

At the end, I test the final filter derived from T100 adaboost and realboost to image taken at the class.

Adaboost T100



There are some faces which are far or wearing glasses can't be recognized. And many false alarm occured from clothes and desk edges.

Realboost T100

With the same iterations, Realboost cnoverges faster and the filter can recoginize more faces.
But still many false alarm occured from clothes of students and from the desk edges. Negative mining with background of class might help reduce false alarms from desk edges, but those from clothes can't be canceled. We can tell this by adjust the threshold,such as following picture turning the threshold to 0.3:

With higher threshold, we detect less faces but have lower false alarms from the desk edges. But false alarms from clothes are still many. The classifer can't tell clothe from human faces. So to improve the result, we should add more cloth images as negative population for negative mining to let our algorithm learn better.