***Boosting***

*Yonglin Zhuo(yonzh457) and Andreea Ioana (andbo713)*

1. ***Plots of the most important Haar features chosen by AdaBoost and some thoughts why these might be important for the classification.***

At first, when we construct the training set and testing, we need to use the Haar features to extract the features of the original data. It should be some particular features are more important.

Secondly,as it known, at the output step of the AdaBoost algorithm, we need to use the formula below to get the final estimate. In our lab, we use each Haar Feature to get each weak learner. The learning rule is so simple. For example, 4 Haar features have chosen and the learning rule for the first weak leaner is that if the first coordinate of one feature vector is less than a specific threshold, then the output should be 1,otherwise, -1. As a consequence, the weight of each weak leaner stands for the importance of each Haar Feature. The higher the weight of weak learner is, the more important the corresponding feature is. Figure1 shows the top 5 important Haar features when we pick up 25 Haar features for one implementation with 2000 training and 1000 testing examples. (5 4 3 11 2) error= 0.1550

5 11 3 31 2 4 63 27 13 95

0.6807 0.3746 0.3369 0.2816 0.2754 0.2658 0.2617 0.2573 0.2492 0.2431

1. ***Plots of some misclassified faces and non-faces that seem hard to classify correctly.***

Nbr of weak learners: 100

Training data = 6000

Testing data = 3000

Error = 0.0623,187 training: 0.0593

1. ***Plot of how the classification accuracy on training and test data depends on the number of weak classifiers. What was the best accuracy you could achieve and how?***

5 error=0.2510

10 error = 0.2280

15 error= 0.1900

20 error = 0.1420

25 error= 0.1550

30 error = 0.1350

35 error = 0.1180

40 error = 0.0930

45 error = 0.1170

50 error = 0.0970

55 error = 0.1100

60 error = 0.1230

65 error = 0.1030

70 error = 0.0960

75 error = 0.0870

80 error = 0.0940

85 error = 0.0990

90 error = 0.0890

95 error = 0.0920

100 error = 0.0820

105 error =0.0890, 0.0235

1. ***Defend your results, are they reasonable? Can we expect perfect results?***

5 error= 0.2700 0.2645

10 error = 0.1730 0.1585

15 error= 0.1630 0.1495

20 error = 0.1620 0.1210

25 error= 0.1440 0.1025

30 error = 0.1310 0.0945

35 error = 0.1330 0.0885

40 error = 0.1240 0.0830

45 error = 0.1150 0.0775

50 error = 0.1060 0.0625

55 error = 0.1020 0.0625

60 error = 0.0960 0.0560

65 error = 0.0940 0.0550

70 error = 0.0950 0.0500

75 error = 0.0980 0.0440

2000,1000

10 error = 0.1930 0.1955

20 error = 0.1320 0.1270

30 error = 0.1160 0.1020

40 error = 0.1130 0.0770

50 error = 0.1040 0.0610

60 error = 0.0940 0.0510

70 error = 0.0900 0.0450

80 error = 0.0850 0.0400

90 error = 0.0800 0.0340

100 error = 0.0770 0.0300

Top\_10 = 6 9 36 8 30 16 20 41 29 14

Nbr\_miss = 77