B657 – Assignment 3

Arun Ram Sankarnarayanan, Raghavendra Nataraj, Ganesh Nagarajan

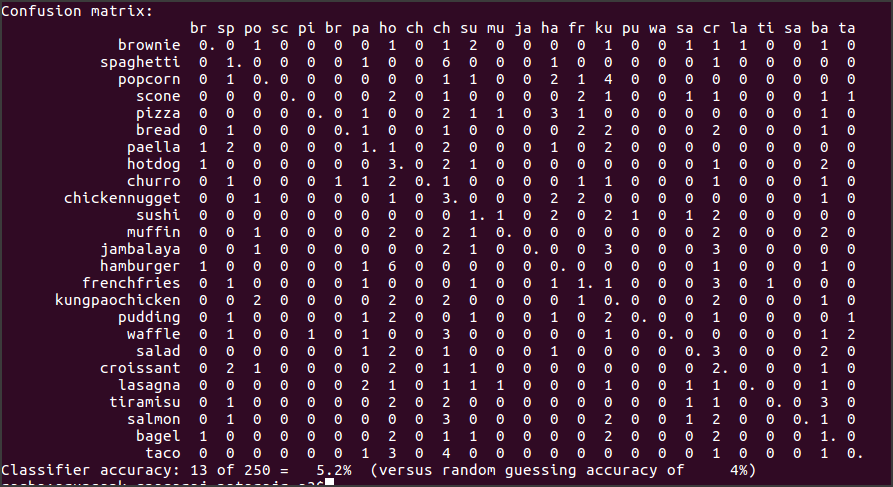
1. Baseline Features

For baseline results, we choose to sample the image to a size of 100. Hence the resulting vector for grayscale is 10000. In case of color images, the sampled image dimension turns out to be 30000.

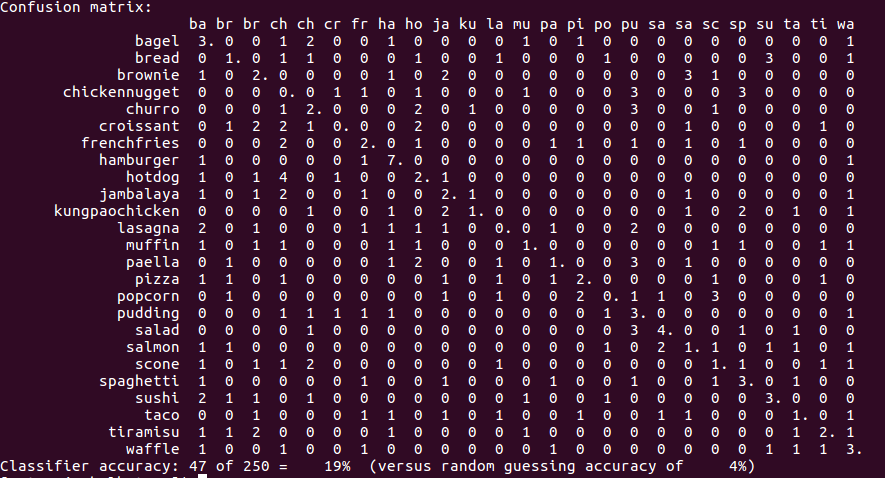
Also for comparison, a 50x50 image is also trained.

Greyscale Confusion matrix

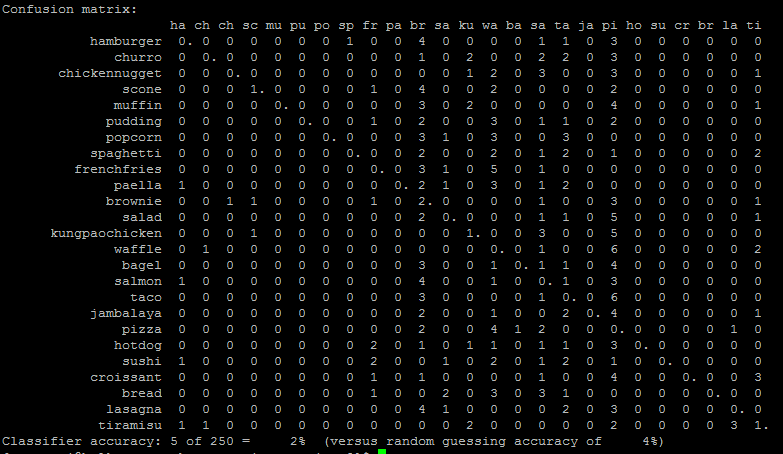
100\*100 greyscale



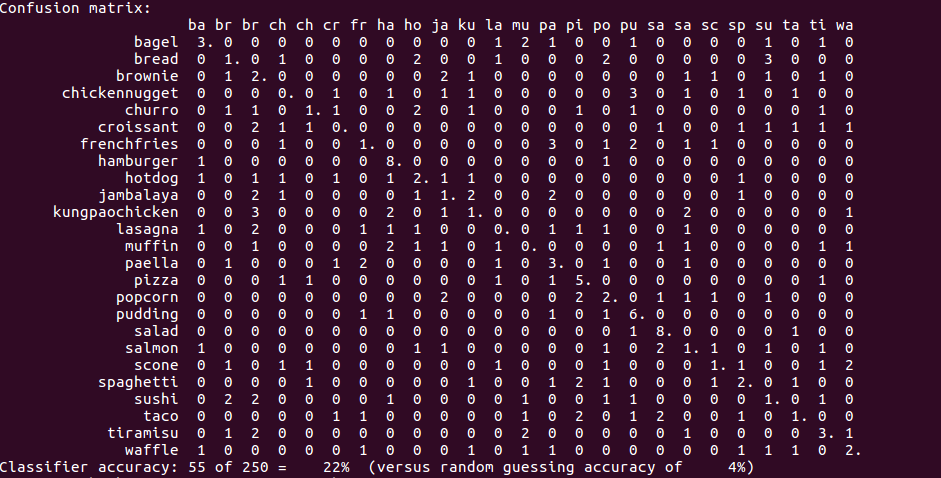
100\*100 color



50\*50 grey scale



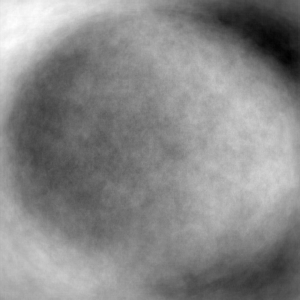
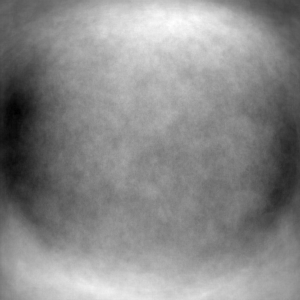
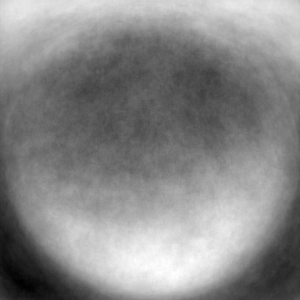
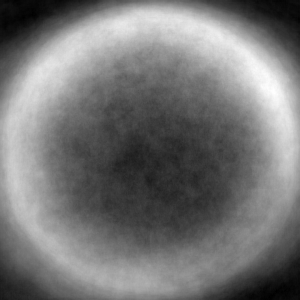
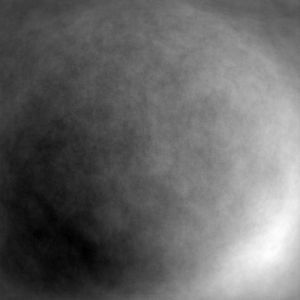
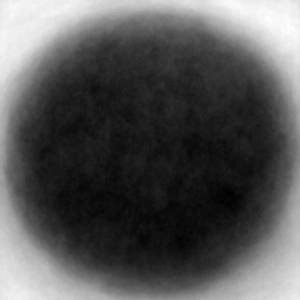
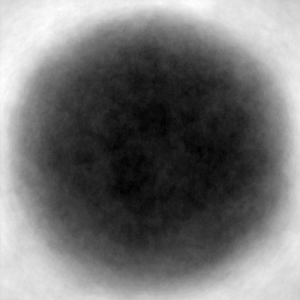
50\*50 color



1. Traditional Methods

Eigen Food:

Initially we resized the images to 100 x 100 and considered up to 200 principal components.



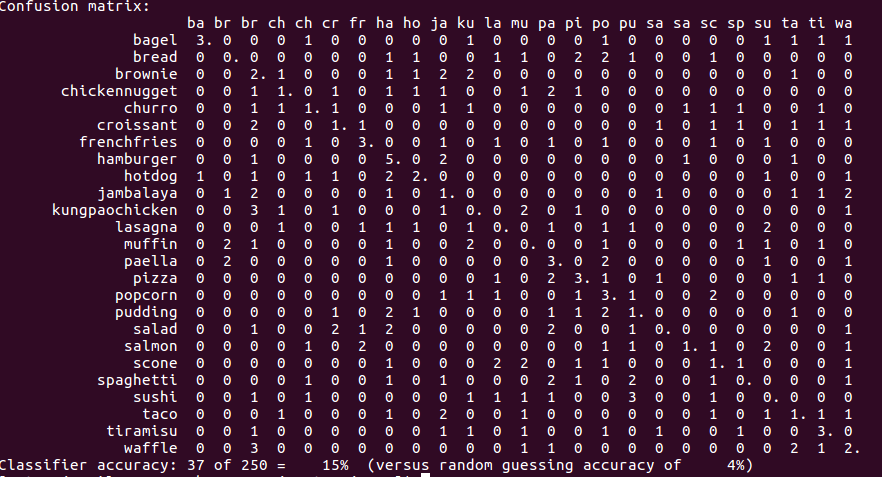
Images of Top 10 Principal Components for a 100 \* 100 image.

From the above experiment, it seems that using first 200 components give the best performance of about 15% accuracy. Also at every time a 100 \* 100 image performs better than a 50 \* 50 image.

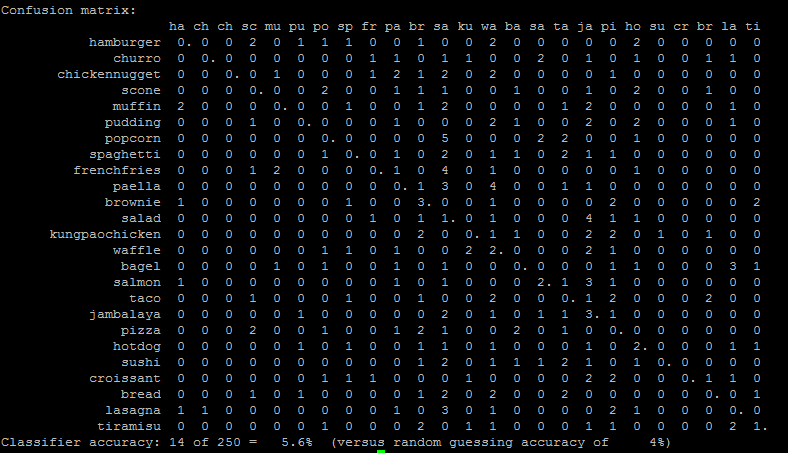
The above graph explains the variability explained by the principal components.

It can be seen that the first principal component gives the maximum variance and this reduces as we move on. It can be seen that there is a considerable difference between the consecutive components.

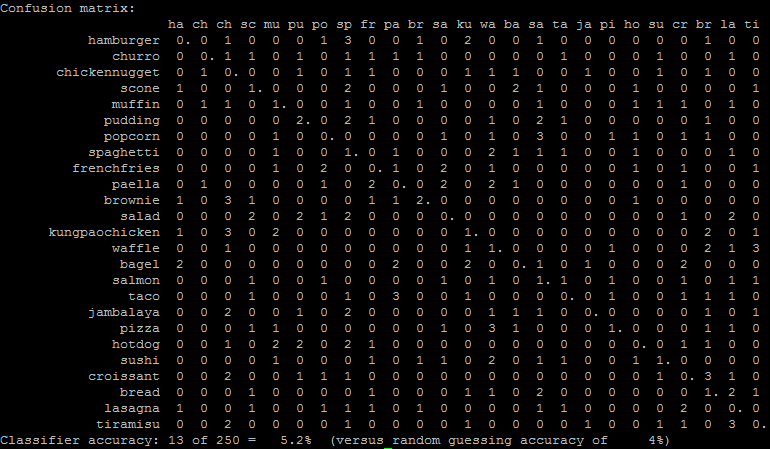
100\*100 200 components:



100, 100 dim

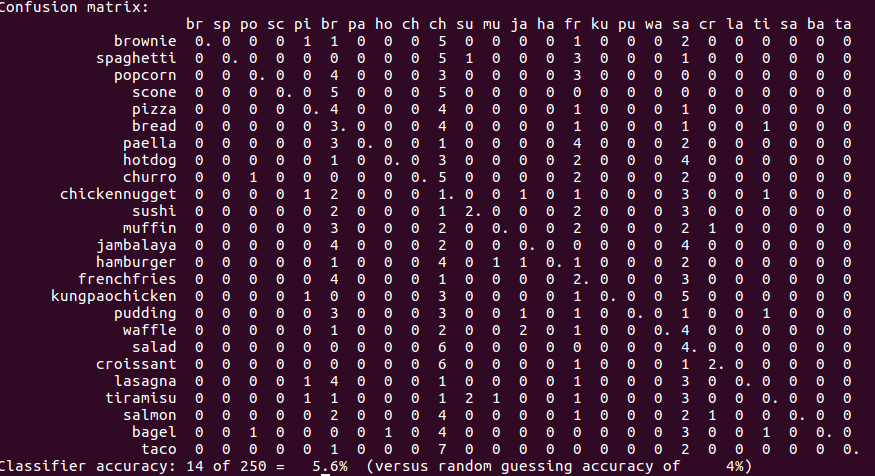


100 -> 100



Haar Like Features:

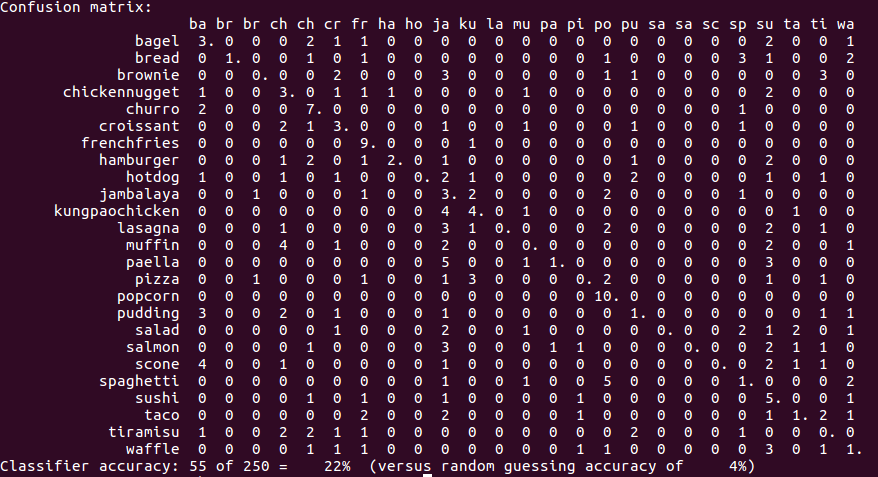
For random features i.e random rectangular regions, we generated a predefined set of rectangular regions. We used these predefined rectangular shapes against the integral and summed up the pixels underneath the rectangle and these are now used as a feature for SVM.



We have experimented with other variations like changing the dimensions of the rectangle, however certain variants result in very bad accuracy. Due to long execution time, we weren’t able to explore with more parameters. We have attached the best outcome from our experiment.

Bag of Words:

Bag of words model uses the SIFT features and these features are clustered by k-means algorithm. We noticed that we got the maximum accuracy at 250 clusers, however it reduced ask reduces. Also we noticed that for 100 iterations of k-means, a 250 cluster took about an hour to run in silo.soic.indiana.edu

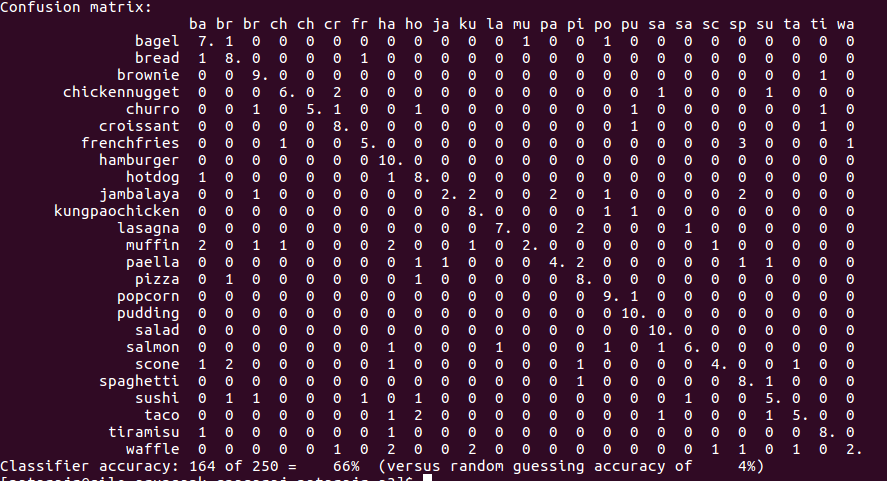


**Comparision among Methods**

We can see that BOW features as of now gives the best seen accuracy of 22% while comparing to the best of 15% of Eigen foods.

Deep Learning

We dumed the layer 21 features of the fast network using the overfeat binary. This was 1000 values. This was fed into the SVM and trained. Similarly test images were also overfeat to dump and we pass that through the SVM model for prediction.



It can be seen that the CNN has the best accuracy of about 66%, the highest among all the methods that we discussed.