

Don Ma  
CS 482 – Fall 2012  
HW4

Method	Error Rate (% of test set incorrectly classified)	Error Rate of K-means label on training set
PCA, K-mean Labels	88.8%	78%
PCA w/o first 4 components, K-mean Labels	42.3%	31%
PCA, Hand labeled	77.3%	-
PCA w/o first 4 components, Hand Labeled	34.2%	-

The error rates were pretty large for every attempted method, which makes me worry that I didn't follow some given instruction, (I accidentally deleted the instructions you wrote on the whiteboard)

My first pass I did the following:

1. Load training set images
2. Get average of training set
3. normalize training set with the average
4. get SVD of the new set
5. Take the first 9 of the U components
6. Find training set weight in the 9 component space
7. K-means cluster the 190 images, in the 9 space, with  $k=10$ , classes
8. get the test set weights in the 9-d subspace
9. using the 10 cluster centers, find the closest for each of the 260 test vectors,

I did a few modifications to try and improve the errors;

As suggested by the paper, I removed the first 4 components, since they seem to be lighting indicators.

(this is a part of the reason I'm worried I didn't do something with the lighting that I was supposed to)..

I also just hand labeled the training set to get the ground truth, instead of the K-means clustering. (which I remember you saying to use).

Over all ignoring the first 4 components of the PCA, and using 5:13, drastically improved the accuracy.

I was also unsure how to find the error rates using K-means, since the ground truth is lost (since the  $k$  clusters are stochastically place). I did my best by finding the median value of each of the 10 clusters as a labeling to find the error rates in the test set.

I did not actually implement the fisherfaces algorithm from the paper, even though it is said to have better results.