SMAI Mini Project

Fall Monsoon 2013

Kshitij Kansal
Lab For Spatial Informatics
IIIT-H
Hyderabad, India
ksitij.kansal@students.iiit.ac.in

Nehal J Wani
Search and Information Retrieval Laboratory
IIIT-H
Hyderabad, India
nehal.wani@students.iiit.ac.in

***Abstract*—Our approaches treat the facial recognition problem as an intrinsically 2D recognition problem. The system functions by projecting face images onto a feature space that spans the significant variations among known face images. The significant features (eigenfaces) are the principal components of the sets of faces. The projection operation characterized an individual face by a weighted sum of the eigenface features, and so to recognize a particular face it is necessary only to compare these weights to those of known individuals.**

Keywords—eigenfaces; recognition; features; eigenvectors;

I. GOAL

The goal of this mini project is to get ourselves familiarized with the ideas of image representation, PCA and LDA, and face recognition and to understand the practical difficulties in developing real-world systems that work with acceptable accuracies.

II. *1.* *INTRODUCTION*

Bio metric recognition, including face recognition, can be carried out in two modes: a) Identification, and b) Verification. In both cases, a few training samples from each class (each person) is collected and stored in a reference database during training. This process is called enrollment. During testing, a test sample is presented, which should be compared against the enrolled samples to decide if they match or not. The matching process differs in the two modes described above. In the identification mode, the test sample is compared against training samples of each of the enrolled persons, and the best match is identified to decide the identity of the person providing the test sample. This is a one-to-many matching process. In the verification mode, the test sample is presented along with a claim of identity. The test sample is matched only against enrollment samples of the claimed identity to decide whether it is a match or not. The process hence verifies whether the identity claim is correct or not. This can be thought of as a one-to-one matching, even though one might compare the test sample against multiple enrollment samples of the claimed person. We find a linear projection of the faces from the high-dimensional image space to a significantly lower dimensional feature space which is insensitive both to variation in lighting direction and facial expression. We choose projection directions that are nearly orthogonal to the within-class scatter, projecting away variations in lighting and facial expression while maintaining discriminability.

III. *2.* *OVERVIEW*

The Principal Component Analysis (PCA) is one of the most successful techniques that have been used in image recognition and compression. PCA is a statistical method under the broad title of factor analysis. The purpose of PCA is to reduce the large dimensionality of the data space (observed variables) to the smaller intrinsic dimensionality of feature space (independent variables), which are needed to describe the data economically. This is the case when there is a strong correlation between observed variables. This has been further described in section 4.1(A).

Classification methods based on KNN utilize a similarity or distance measure between images. The basic idea is that an incoming samples, will be classified according to the classes assigned to the training data that are k nearest neighbors. This classification method is popular because it is simple, intuitive, and easy to implement. Furthermore, it has shown to perform well in our case and the results are shown in the section 4.1(B).

In machine learning, support vector machines (SVMs) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. Support vector machines have met with significant success in numerous real-world learning tasks. However, like most machine learning algorithms, they are generally applied using a randomly selected training set classified in advance. SVM: SVM method produces better accuracies as compared to KNN. SVM is better if we have non-linear decision boundary. KNN is prone to over fitting. This is explained in section 4.1(B).

A receiver operating characteristic (ROC) curve is commonly used to measure the accuracy of a diagnostic test. It is a plot of the true positive fraction (sensitivity) against the false positive fraction (1-specificity) for increasingly stringent positivity criterion. Bias can occur in estimation of an ROC curve if only some of the tested samples are selected for verification and if analysis is restricted only to the verified cases. This bias is known as verification bias.

Data validation is the process of ensuring that a program operates on clean, correct and useful data. It uses routines, often called "validation rules" or "check routines", that check for correctness, meaningfulness, and security of data that are input to the system. Here two validation techniques have been used, **4 - cross validation** and **hold one out validation.**

3. Datasets

The Yale Clipped Face Database (size 88.4MB) contains 2587 gray scale images in Portable gray Map (PGM) format of 38 individuals. There are 67 images per subject, one per different facial expression or configuration: center-light, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprise, wink and in different illumination and lighting conditions. Each image is 168 x 192 pixels in size. From all the 67 given images of each person, only 20 are selected for having the Azimuth and Elevation angles between -25 and +25 degrees. The user can then resize them to a small scale of let us say, 80X80.

CMU PIE Database. It contains images of 68 person under 13 different poses, 43 different illumination conditions, and with 4 different expressions. The mat file containing this was available which contained the images and its corresponding labels in a structure. Each image is about 32 x 32 pixels in size. The user can resize them to some different scale.

IIIT-H SMAI 2013 Student Database. The database has generated by data collected over a period of three months in various backgrounds, surroundings and lighting. The database is able to capture the changes in faces along the time like growth of hairs, beard etc. It is also able to capture the changes in expression of a student during different phases of the class. It captures the natural expressions of the person thus providing a data more close to real world problems.

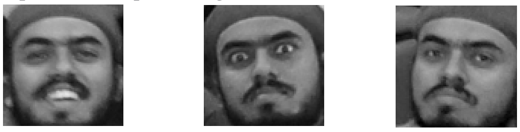

Fig. 1. Real time Capture of a student's expression

4. *Experiments and Results*

In the subsections below, we have described in detail the experiments conducted and the corresponding results obtained. These results are then supported by some graphs for better comparisons and understanding.

4.1 Experiments performed using classifiers.

A. Feature Extraction

For extracting the feature of the images, a technique called Principal Components Analyses(PCA) was used. The basic idea behind PCA is that the data (in our case, image) in very high dimension space(order of thousands) can be projected in some small dimension space without loss of much information. The technique basically tries to extract eigen faces from the training data. These eigen faces, form the new required reduced space. For calculating these eigen faces, another technique proposed by [1] has been used. We basically find the eigen vectors of the matrix A^TA, where A is the matrix of the mean subtracted images stored columwise.

$$A^T AV = \lambda V$$

$$\left(AA^T\right) AV = \lambda AV$$

Assumption: N^2XN^2 is much much bigger than number of images.

After getting the eigen values of A^TA, we pre multiply them by A to get the eigen values of AA^T, which is nothing but the co-variance matrix. This step is done to reduce number of computations as AA^T is N^2XN^2 dimension and finding its eigen vectors directly is computationally heavy.

The eigen vectors are sorted in decreasing order on the basis of their eigen values, and the top *k* are taken to define the new reduced dimension space. The weight vectors are now calculated used the following equation:

$$W = V_i\left(\Im - \wp\right)$$

where, $\Im$ is original image and $\wp$ is mean of training images.

B. Classifiers

SVM Classifier: The SVM classifier constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.

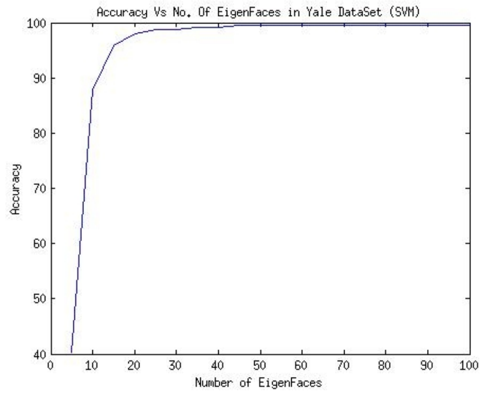

Fig. 2. SVM on Yale Dataset

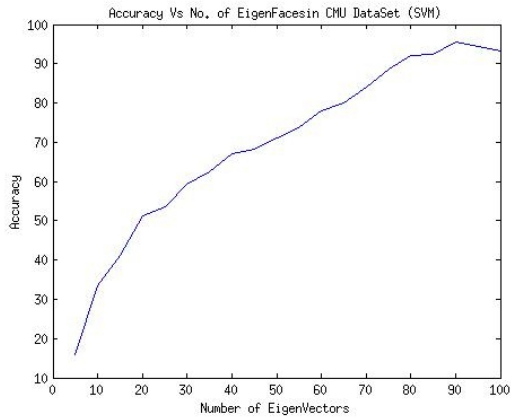

Fig. 3. SVM on CMU Dataset

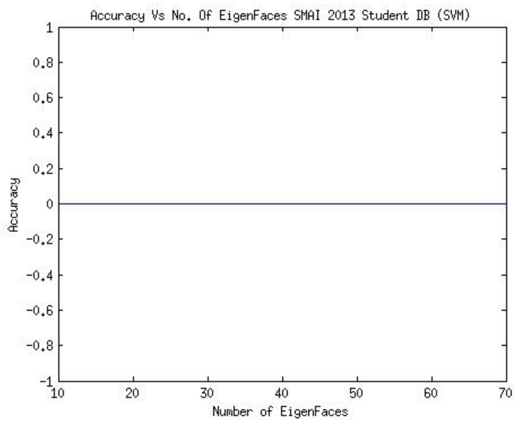

Fig. 4. SVM on SMAI 2013 Student Dataset

Analysis:

Fig 2, 3 & 4 depict the accuracies of the SVM classifier that was run after extracting features using PCA on the three Datasets. From the above plots, it can be interpreted that as the count of eigen faces increases, the accuracy increases logarithmically. The steep increase in the accuracy in the starting shows us that most of the information is stored in the first few eigen vectors. The variations in the above three plots can be explained as follows: In the Yale Dataset, the images were taken in man made conditions by fixing the illumination, lighting and many other conditions. So the data is linearly separable in some high dimension and hence good result in SVM. The CMU Dataset contains little more variations as compared to Yale Dataset and hence a little distorted graph. On the other hand, the SMAI 2013 Student Dataset is a real time capture of images without any man imposed restrictions. The images obtained are not even separable in higher dimension space and use of hold one out method instead of 4 cross validation, which is the reason behind the zero accuracy.

K-Nearest Neighbor: The K-nearest neighbors algorithm (k-NN) is a non-parametric method for classification that predicts objects' "values" or class memberships based on the k closest training examples in the feature space. k-NN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification.

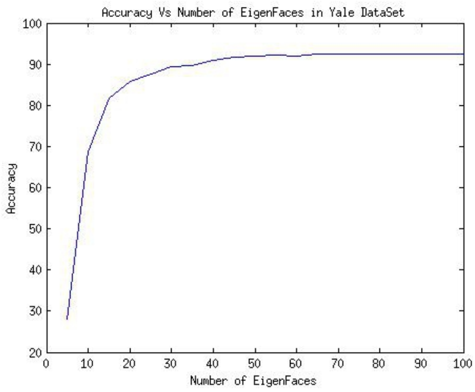

Fig. 5. KNN on Yale Dataset (k=5)

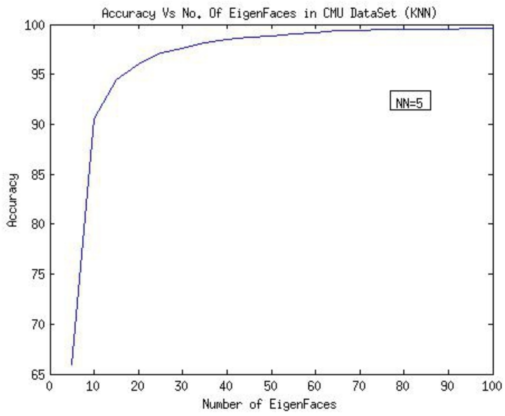

Fig. 6. KNN on CMU Dataset (k=5)

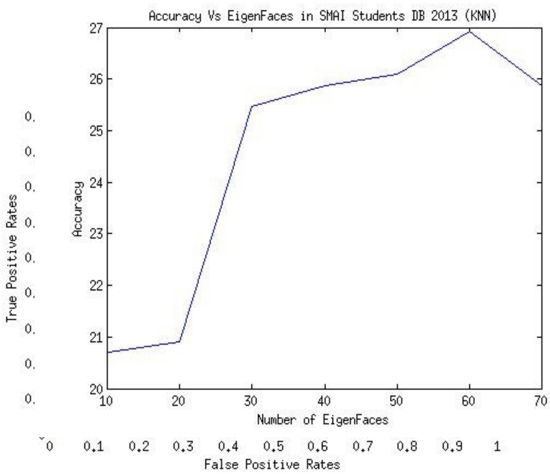

Fig. 7. KNN on SMAI 2013 Student Dataset(k=5)

Analysis:

Fig 5, 6, & 7 denote the accuracies of k-NN classifier on the three Datasets as the number of eigen faces is increased, keeping the value of *k* fixed. The first two Datasets with manually imposed conditions show the results according to our expectations. However, when compared with the SVM classifier, k-NN performs better in the case of CMU and the results are comparable in the case of Yale. When the same classification method is applied to SMAI 2013 Student Dataset, the graph shows irregularity which can be attributed to the fact that the data has been taken in real time without man-imposed restrictions.

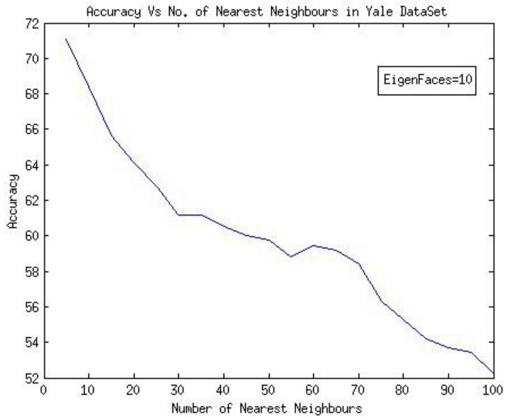

Fig. 8. KNN on Yale Dataset

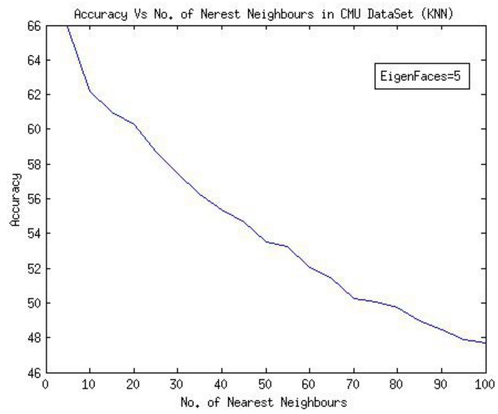

Fig. 9. KNN on CMU Dataset

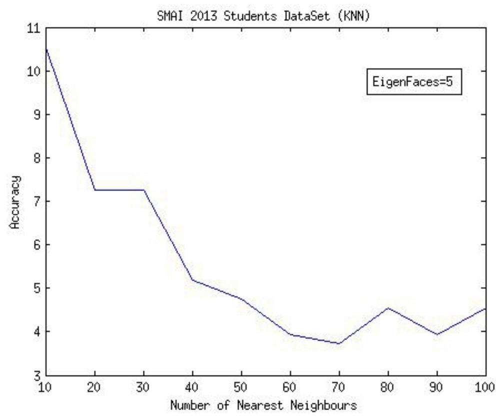

Fig. 10. KNN on SMAI 2013 Student Dataset

Analysis:

The graphs presented in Fig 8, 9 & 10 highlight the fact that as we increase the value of k for calculating the class of the testing data, there is a gradual decrease in accuracy, which is due to high variance in the Dataset.

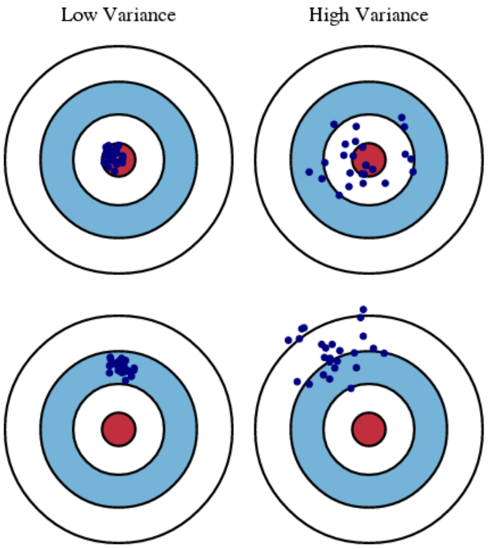

Fig. 11. Graphical illustration of variance.

C. *Excluding Initial Eigen Vectors*

It is observed that if we exclude some initial eigenfaces, the accuracy increases dramatically, The reason behind this is that the randomness is very high in top few eigenfaces. These top eigenfaces capture the variations due to surroundings and thus by eliminating them we will be eliminating the effect of surroundings. The results of this experiment is captures in Fig. 12&13.

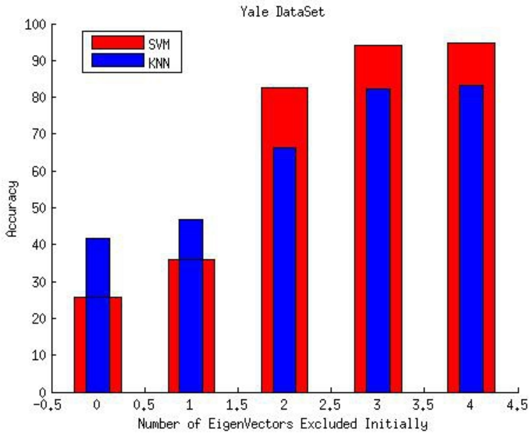

Fig. 12. Excluding Initial few eigenfaces. (Yale)

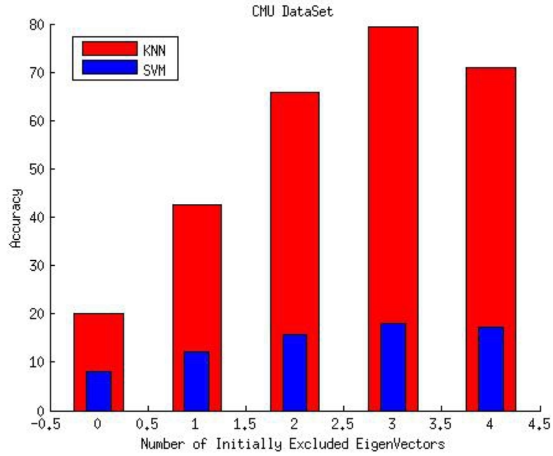

Fig. 13. Excluding Initial few eigenfaces. (CMU)

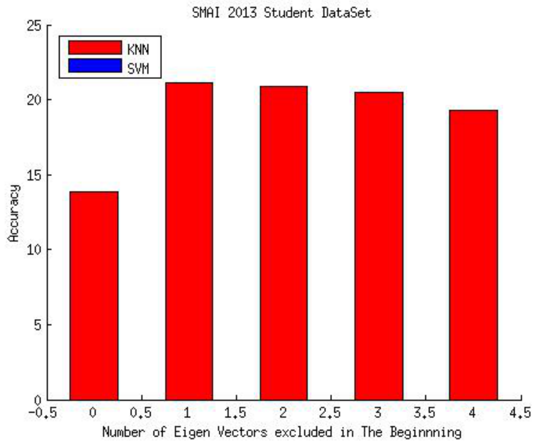

Fig. 14. Excluding Initial few eigenfaces. (SMAI 2013 Student Dataset)

D. *Receiver operating characteristic*

A Receiver Operating Characteristic (ROC), or simply ROC curve, is a graphical plot which illustrates the performance of a binary classifier system as its discrimination threshold is varied. It is created by plotting the fraction of true positives out of the total actual positives (TPR = true positive rate) vs. the fraction of false positives out of the total actual negatives (FPR = false positive rate), at various threshold settings. ROC plots for the Yale and CMU Datasets are plotted in the Fig. 15&16.

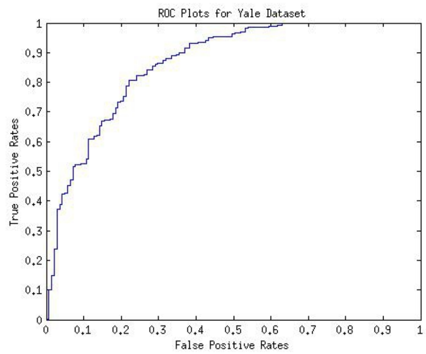

Fig. 15. ROC plot for Yale Dataset

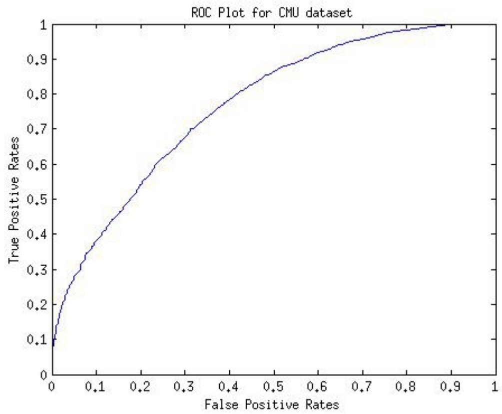

Fig. 16. ROC plot for CMU Dataset

E. *Image Reconstruction*

Once we have the reduced eigen faces, we can reconstruct any image using them. Upon increasing the number of eigen faces, the reconstructed image is more close to the original image.

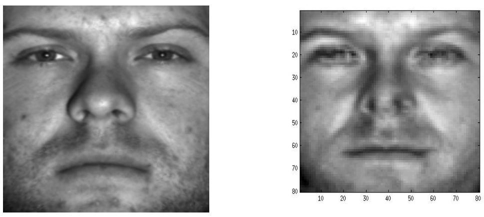

Fig. 17. Original and Reconstructed image (no of eigen faces = 30)

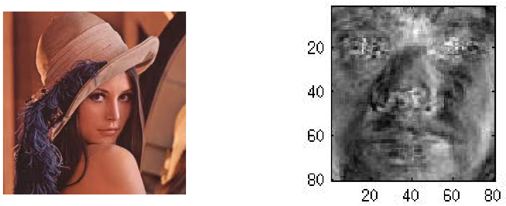

Fig. 18. Original and reconstructed image

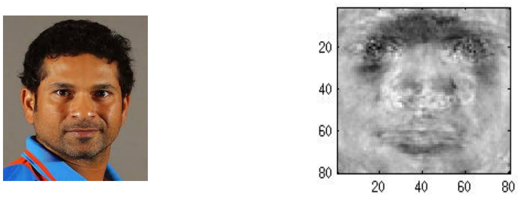

Fig. 19. Original and reconstructed image

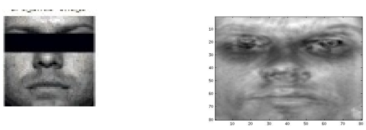

Fig. 20. Eyes covered and the reconstructed image

Fig. 21. Original and Reconstructed image (no of eigen faces = 30)

Analysis:

The Fig. 17 is the reconstruction of the image from the training set itself. Results for reconstructing face images not in the training set is shown in Fig. 18 & 19. We can see that Fig. 18 which has lot more variations in surroundings(Hat, shoulder and side ways view of face) is reconstructed badly as compared to the Fig 19. which is straight face image with less background variations.

This whole technique can be used to extract certain parts of face when they are not visible. Fig. 20. Also the results for reconstruction from a non-face image is shown in Fig 21.

5. *References*

[1].“Eigenfaces for Recoganition”- Mark Turk and Alex Pentland http://www.cs.ucsb.edu/~mturk/Papers/jcn.pdf

[2].“Eigen Faces vs. Fisherfaces: Recoganition Using Class specific Linear Projection” - Peter N. Belhumeur, Jo~ao P. Hespanha