Comparative Study of PCA, KPCA, KFA and LDA Algorithms for Face Recognition

Nitin S Prakash[1], Prathviraj Shetty[2], Kiran Kedilaya[3] , Nithesh[4], Sunil B N[5]

[1,2,3,4,5] *Dept of Computer Science and Engineering, Sahyadri College of Engineering and Management, India*

Abstract - *Face recognition is the process of identifying one or more faces in images or video clips by analysing and comparing known patterns. One of the ways to do this is by comparing selected facial features from the image and a facial database. Algorithms typically extract facial features and compare them to a database to find the best match. This has wide range of applications which includes Security, Criminal Identification, Payments, Healthcare, Marketing and so on. There are plenty of algorithms devised and used for face recognition, each having its own benefits and shortcomings. For any face recognition algorithm, the two key stages are extraction of features and their classification. In this paper we focus on comparison between the performances of PCA, KPCA, KFA and LDA algorithms against AT&T, Yale and UMIST datasets.*

Key Words: PCA, KPCA, KFA, LDA, UMIST, AT&T, Yale, face recognition.

1.INTRODUCTION

As humans, recognising faces is no big deal to us. But for a computer which understands only the binary language consisting of 0 and 1, it is a huge challenge to identify human faces. Humans of age 2-3 years old can recognise and distinguish between known faces distinctively. Our knowledge regarding the process of recognising faces happening inside a human brain is limited. An approach based on geometric facial features is the most innate method for recognising faces. One of the earliest approaches for face recognition was based on euclidian distance between the feature vectors of a test image and reference image. This method, even though was unaffected by change in lighting conditions, had a major setback due to the complicated task of accurately positioning marker points.

Face recognition differs substantially from traditional pattern recognition problems as the pattern to be recognised doesn't alter. In object recognition, the shapes of the objects change while in face recognition objects are recognised with the same rudimentary shape differing in colour and form. Although plenty of algorithms prevail in this domain, each having their respective utility and hiccups.

While doing a comparative analysis of algorithms, it is recommended to run them against common and standard datasets. This helps in easier and efficient comparison between their performances.

In pattern recognition and in image processing, feature extraction based dimensionality reduction plays an important role in the relative areas. Feature extraction simplifies the amount of resources required to describe a large set of data accurately for classification and clustering [1]. Transforming the input data into the set of features is called feature extraction. Due to the complex variations of illumination, expression, angle of view and rotation, etc., it is difficult to describe the facial features through a single algorithm. Therefore, most of the current researches on face recognition focus on the recognition problems under restricted conditions. A common face recognition system consists of preprocessing, feature extraction/selection, and recognition.

1.1 Principal Component Analysis (PCA)

PCA is a statistical methodology used to minimise the large dimensionality of the data to manageable congenital dimensionality. This is essential to depict the data in a cost effective manner. PCA is an approach that acts in the linear space hence applications having linear prototypes such as signal processing, image processing, system and control their, communications are compatible. Compact principal components of the feature space are obtained by PCA after compressing the huge single dimensional vector of pixels generated from 2-Dimensional facial image. This is called eigenspace projection. By distinguishing the eigenvectors of the covariance matrix acquired from a set of vectors the eigenspace is determined.

1.2 Kernel Principal Component Analysis (KPCA)

A linear transformation is not suitable for capturing the nonlinear structures of the data. In order to represent the nonlinear structure, the Kernel PCA (KPCA) has been formulated. In KPCA, the computational cost depends on the sample size. When the sample size is very large, it is impractical to compute the principal components via a direct eigenvalue decomposition. Through adopting a polynomial kernel, the principal components can be computed within the space spanned by high- order correlations of input pixels making up a facial image, resulting in an improved performance [2].

The main idea of KPCA is to project the input data from the linear space into the nonlinear space and then implement PCA in the nonlinear feature space for feature extraction. In KPCA, the nonlinearity is firstly mapping the data into another space using a nonlinear map, and then, PCA is implemented using the mapped examples [3]. The mapping and the space are determined implicitly by the choice of a kernel function which computes the dot product between two input examples mapped into feature space via kernel function. If kernel function is a positive definite kernel, then there exists a map into a dot product space.

KPCA-based feature extraction needs to store the original sample information owing to computing the kernel matrix, which leads to a huge store and a high computing consuming [3]. It is feasible to improve the performance of KPCA with sparse analysis and kernel optimisation. We reduce the training samples with sparse analysis and then optimise kernel structure with the reduced training samples[3].

1.3 Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) has been successfully applied to face recognition which is based on a linear projection from the image space to a low dimensional space by maximising the between class scatter and minimising the within-class scatter [4]. LDA allows objective evaluation of the significance of visual information in different features of the face for identifying the human face. The LDA also provides us with a small set of features that carry the most relevant information for classification purposes.

LDA method overcomes the limitation of Principle Component Analysis method by applying the linear discriminant criterion. This criterion tries to maximise the ratio of determinant of the between-class scatter matrix of the projected samples to the determinant of the within-class scatter matrix of the projected samples [4]. Linear discriminant groups the images of the same class and separate images of different classes. Here to identify an input test image, the projected test image is compared to each projected training, and the test image is identified as the closest training image [4].

Linear discriminant analysis is primarily used here to reduce the number of features to a more manageable number before classification. Each of the new dimensions is a linear combination of pixel values, which form a template [5]. The linear combinations obtained using Fisher's linear discriminant are called Fisher faces, while those obtained using the related principal component analysis are called eigenfaces.

There are two main phases: training and classification. In the training phase, a fisher space is established from the training samples and the training faces are projected onto the same subspace. The optimal projection (transformation) can be readily computed by applying the Eigen decomposition on the scatter matrices. In the classification phase, an input face is projected into the fisher space and classified using the Mahalanobis Cosine distance as a similarity measure [6].

1.4 Kernel Fischer Discriminant Analysis (KFDA)

KFDA (also known as KFA - Kernel Fisher Analysis) is a method that is developed from Linear Discriminant Analysis. It is a non-linear classification technique based on Fisher's discriminant. The main ingredient is the kernel trick which allows the efficient computation of Fisher discriminant in feature space [7]. The linear classification in feature space corresponds to a (powerful) non-linear decision function in input space.

LDA is used to maximise the separation of data in two different classes and minimising the distance between data in the same class. The combination of LDA with kernel function cause nonlinear transformation. The benefit of KFDA is that we can process a lot of data with a high dimensional vector. A high dimensional data is reduced into the small one and then it performing a new projection vectors. The objective is that when we have to deal with a lot of data and dimensions, we can classify the data much better than only using LDA.

2. DATASETS

It is recommended to pitch algorithms against a standard database to get a benchmark for their comparative analysis. We make use of the following standard datasets for the experimental phase: AT&T "The database of Faces" (formerly TheORL Database of Faces), The Yale Face Database and The UMIST Face Database.

2.1 AT&T The Database of Faces

The AT&T face database contains a total of 400 images of 40 unique subjects with 10 images per subject. The images are greyscale and are all stored in PGM format. Each image has dimensions of 92 X 112 pixels. The images are taken from the frontal view and some of the images have a slight tilt of the head in order to introduce variations.

2.2 The Sheffield (Previously UMIST) Face Database

The Sheffield (Previously UMIST) database contains a total of 480 images of 20 unique subjects with 24 images per subject with differences in race, gender, appearance. Each individual is shown in a range of poses ranging from right profile to left profile. The files are all in PGM format, having dimensions 220 x 220 pixels with 256-bit grey-scale.

2.3 The Yale Face Database

The Yale face database contains a total of 165 images of 15 subjects with 11 images per subject. The images are all in greyscale and are stored in gif format. Each imagehas a dimension of 320 X 243 pixels. The images contain the following varying conditions: centrally lighted, with glasses, happy, left light, without glasses, normal, right light, sad, sleepy, surprised and wink. All these various conditions provide an ample amount of features to be detected and used for classifications to check various aspects of the algorithm to be tested.

3 EXPERIMENTAL RESULTS

Experimental results are all taken in terms of first rank recognition rate.

3.1 Algorithm Specific

Algorithm specific results help us determine the performance of individual algorithms when pitched against various datasets with varying training ratios. This gives better insights into how exactly they perform under the various postures and illumination offered by the datasets. We have taken the data sequentially for training.

3.1.1 Principal Component Analysis (PCA)

From the PCA algorithm results we observe that for lower training ratios the efficiency is greater for AT&T dataset indicating that the algorithm needs less training to recognise frontal region of the human face. As the training ratio increases, the result on Yale face dataset shows better recognition rate indicating an improved face recognition rate for varying facial features and making the system more robust

Table 1: Face recognition rates of PCA

Training Set Percentage	DATASETS		
	AT&T	Yale	UMIST
30%	78.21	60.83	57.19
40%	80.00	72.38	50.36
50%	77.50	66.67	48.75
60%	**80.63**	72.00	54.00
70%	79.17	80.00	54.29

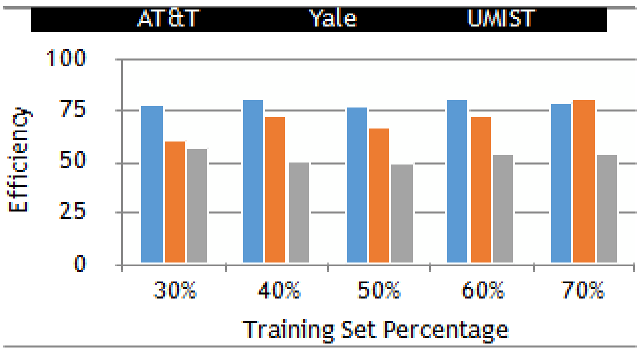

Chart 1: Face recognition rates of PCA

3.1.2 Kernel Principal Component Analysis (KPCA)

From the KPCA algorithm results we observe that for lower training ratios the efficiency is greater for AT&T dataset indicating that the algorithm needs less training to recognise frontal region of the human face. As the training ratio increases, the result on Yale face dataset shows better recognition rate indicating an improved face recognition rate for varying facial features and making the system more robust. The overall efficiency is low due to the fact that the algorithm is designed for non linear data and the data we use is of linear nature.

Table 2: Face recognition rates of KPCA

Training Set Percentage	DATASETS		
	AT&T	Yale	UMIST
30%	49.29	21.67	28.06
40%	61.67	34.29	22.67
50%	52.00	56.67	18.46
60%	58.75	**69.33**	20.45
70%	50.83	58.33	15.00

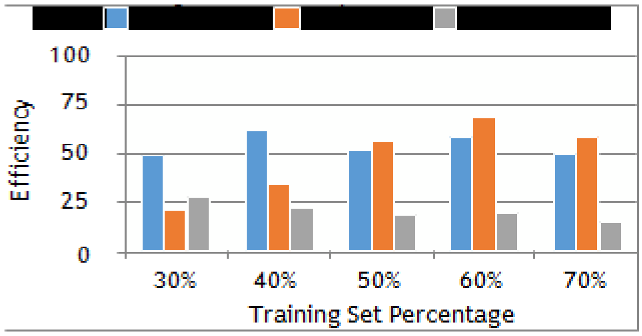

Chart 2: Face recognition rates of KPCA

3.1.3 Kernel Fischer Discriminant Analysis (KFDA)

The KDFA algorithm showcases greater recognition rates for lower training ratios of AT&T dataset as compared to the rest of datasets. With the increase in training ratio, it shows comparable efficiency in recognition with Yale dataset. We observe a steep decrease in recognition rate when it comes to UMIST dataset due to the nature of UMIST dataset and majorly due to the fact that the data is read sequentially. KFDA also shows the highest recognition rates as compares to other algorithms under consideration with a maximum recognition rate of 96.67% on AT&T and Yale datasets for training ratio of 70%.

Table 3: Face recognition results of KFDA

Training Set Percentage	DATASETS		
	AT&T	Yale	UMIST
30%	87.51	81.67	58.89
40%	90.00	84.76	65.33
50%	88.50	87.78	65.38
60%	94.38	85.33	70.91
70%	**96.67**	**96.67**	79.38

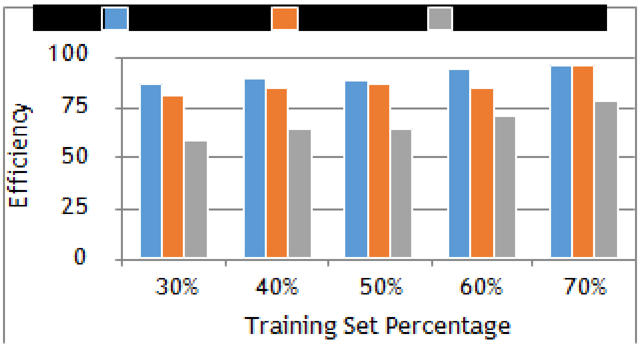

Chart 3: Face recognition results of KFDA

3.1.4 Linear Discriminant Analysis (LDA)

When it comes to LDA, we observe a steady increase in recognition rates for all the three datasets. Yale dataset shows maximum increase in efficiency followed by UMIST and AT&T. Even though AT&T shows least amount of increase in recognition rates, it still has maximum recognition rate up until 60% training ratio which is the dwarfed by recognition rate for Yale dataset at 96.67%.

Table 4: Face recognition rate of LDA

Training Set Percentage	DATASETS		
	AT&T	Yale	UMIST
30%	86.07	80.00	60.83
40%	90.00	80.00	63.00
50%	91.50	90.00	63.69
60%	92.50	88.00	69.09
70%	95.00	**96.67**	71.88

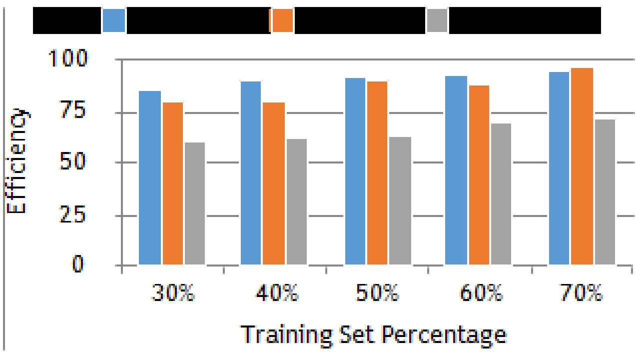

Chart 4: Face recognition rate of LDA

3.2 Dataset Specific

Dataset specific results showcase the comparative efficiency of algorithms with variations in training ratios. This helps determine the type of algorithm to be used under a certain given scenario where the illumination and the system training scenarios are known.

3.2.1 AT&T The Database of Faces

From the results with AT&T Dataset, we observe that the KFA algorithm has obtained the highest recognition rate as compared to all other algorithms hence proving to be the best method to be used for cases where the frontal view of face with slight tilt is the sole condition for the system environment. When the training is limited or very less, both LDA and KFA have comparable recognition rates and can be used interchangeably. The recognition rates of KPCA is dwarfed as compared to KFA due to the use of 'Kernel Trick' in KFA.

Table - 5: Face recognition results of AT&T database

Training Set Percentage	Algorithms			
	PCA	KPCA	KFA	LDA
30%	78.21	49.29	87.51	86.07
40%	80.00	61.67	90.00	90.00
50%	77.50	52.00	88.50	91.50
60%	80.63	58.75	94.38	92.50
70%	79.17	50.83	**96.67**	95.00

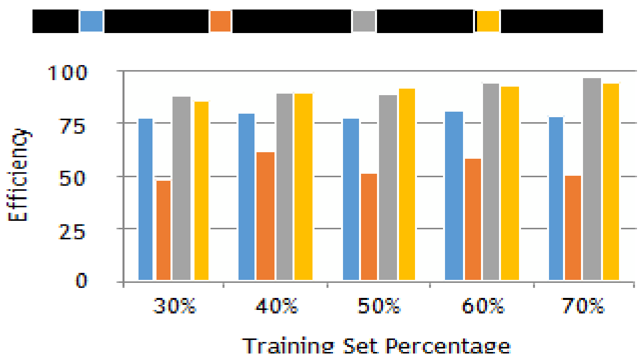

Chart 5: Face recognition results of AT&T Dataset

3.2.2 The Yale Face Database

The Yale Dataset yields out best results with KFA and LDA algorithms hence proving to be the best method to be used for cases where the facial expressions and the direction of illumination varies for each subjects. This is a more robust system environment as compared to the system environment of AT&T Dataset.

Table - 6: Face recognition results of Yale database

Training Set Percentage	Algorithms			
	PCA	KPCA	KFA	LDA
30%	60.83	21.67	81.67	80.00
40%	72.38	34.29	84.76	80.00
50%	66.67	56.67	87.78	90.00
60%	72.00	69.33	85.33	88.00
70%	80.00	58.33	**96.67**	96.67

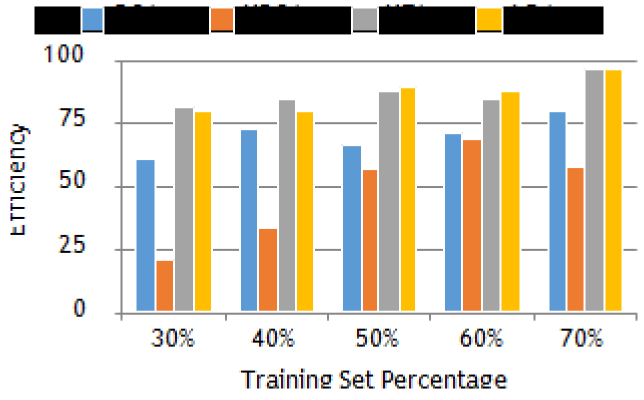

Chart 6: Face recognition results of Yale dataset

3.2.3 The Sheffield Face Database (UMIST)

The Sheffield dataset is tricky dataset when it comes to testing any algorithm with sequential training due to the fact that the subjects' faces start with right profile and slowly rotate by a constant angle until the right profile. The training and testing images have lesser similarities as compared to other datasets and this results in a lower recognition rates when compared to the datasets used here.

Table - 7: Face recognition results of UMIST database

Training Set Percentage	Algorithms			
	PCA	KPCA	KFA	LDA
30%	57.19	28.06	58.89	60.83
40%	50.36	22.67	65.33	63.00
50%	48.75	18.46	65.38	63.69
60%	54.00	20.45	70.91	69.09
70%	54.29	15.00	**79.38**	71.88

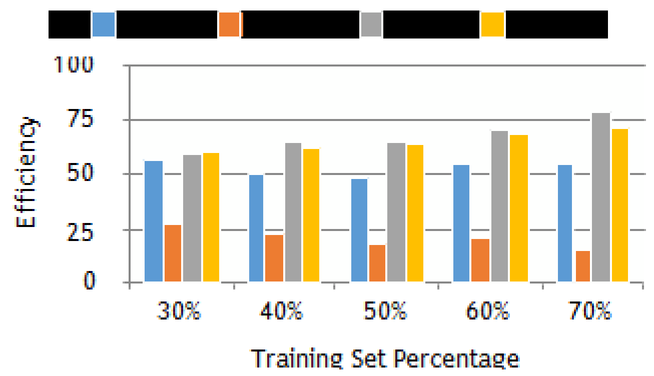

Chart 7: Face recognition results of UMIST dataset

3. CONCLUSIONS

From the various results obtained we observe a trend in the recognition rates of algorithms. We can generally say that as the training ratio increases, we get a increase in first rank recognition rate due to the system being able to identify better features more appropriate to the classification of images. Since the datasets used have a linear data, we observe a steep decline in the recognition rates of nonlinear algorithms as compared to linear algorithms.

KFA algorithm has emerged as the best algorithm among the ones chosen for this comparative study due to the fact that it functions efficiently in both linear and nonlinear image subspace, majorly due to the use of 'Kernel Trick'.

The results of The Sheffield dataset could be improved by using non sequential training methods or by improving the sequential training by taking alternate images for training and testing.

REFERENCES

[1] Li, Jun-Bao, Shu-Chuan Chu, and Jeng-Shyang Pan. "Kernel Principal Component Analysis (KPCA)-Based Face Recognition." Kernel Learning Algorithms for Face Recognition. Springer, New York, NY, 2014. 71-99.

[2] Kim, Kwang In, Keechul Jung, and Hang Joon Kim. "Face recognition using kernel principal component analysis." IEEE signal processing letters 9.2 (2002): 40-42.

[3] Li, Jun-Bao, Shu-Chuan Chu, and Jeng-Shyang Pan. Kernel Learning Algorithms for Face Recognition. Springer, 2016.

[4] Bhattacharyya, Suman Kumar, and Kumar Rahul. "Face recognition by linear discriminant analysis." International Journal of Communication Network Security 2.2 (2013): 31-35.

[5] Premaratne, Prashan. Human computer interaction using hand gestures. Springer Science & Business Media, 2014.

[6] Chelali, Fatma Zohra, A. Djeradi, and R. Djeradi. "Linear discriminant analysis for face recognition." Multimedia Computing and Systems, 2009. ICMCS'09. International Conference on. IEEE, 2009.

[7] Mika, Sebastian, et al. "Fisher discriminant analysis with kernels." Neural networks for signal processing IX, 1999. Proceedings of the 1999 IEEE signal processing society workshop.. Ieee, 1999.