A Comparative Study of FP Principal Component Analysis (PCA) Methods SubXPCA and Others

A Dissertation submitted to the University of Hyderabad

in partial fulfillment of the degree of

Master of Technology

in

Artificial Intelligence

By

Yeduguru Chaitanya

09MCMI26

Department of Computer and Information Sciences

University of Hyderabad

Gachibowli, Hyderabad - 500 046

Andhra Pradesh, India

CERTIFICATE

This is to certify that the dissertation entitled "**A Comparative Study of FP Principal Component Analysis (PCA) Methods SubXPCA and Others**" submitted by **Yeduguru Chaitanya**, bearing Reg. No. 09MCMI26, in partial fulfillment of the requirements for the award of Master of Technology in Artificial Intelligence is a bonafide work carried out by her under my supervision and guidance.

The dissertation has not been submitted previously in part or in full to this or any other University or Institution for the award of any degree or diploma.

Dr. Atul Negi
Project Supervisor,
Department of CIS,
University of Hyderabad.

Head,
Department of CIS,
University of Hyderabad.

Dean,
School of MCIS,
University of Hyderabad.

DECLARATION

I Yeduguru Chaitanya hereby declare that this Dissertation entitled "**A Comparative Study of FP Principal Component Analysis (PCA) Methods SubXPCA and Others**" submitted by me under the guidance and supervision of **Dr. Atul Negi** is a bonafide work. I also declare that it has not been submitted previously in part or in full to this University or other University or Institution for the award of any degree or diploma.

Date: (Yeduguru Chaitanya)

Place: Reg. No. 09MCMI26

To,

My Parents and Friends

Acknowledgments

I would like to express my sincere gratitude to **Dr. Atul Negi**, my project supervisor, for valuable suggestions and keen interest through out the progress of my course of research.

At the outset, i would like to thank **The University of Hyderabad** for providing all the necessary resources for the successful completion of my course work.

At last, but not the least i thank my classmates and other students of DCIS for their physical and moral support.

With Sincere Regards,

Yeduguru Chaitanya.

Abstract

Principal Component Analysis (PCA) is one of the well-known techniques for dimensionality reduction with numerous and diverse applications.Despite the popularity of PCA, its major drawbacks are seen as low selectivity of local features and poor scalability to high dimensional data due to large computational complexity.To address these problems, block based PCA methods (FP-PCA) were proposed.

In this work, we study the comparative performance of various PCA approaches for dimensionality reduction with face recognition as the specific topic. It aims to observe how FP-PCA approaches perform with respect to dimensionality reduction. As an objective criterion of image quality, PSNR was used for the comparative experiments. Images were reconstructed from a variable number of PCs and their quality using PSNR was observed. SubXPCA is a special case of FP-PCA approach with a 2 stage approach. Different versions of SubXPCA are experimented with by choosing components from 1st stage for finding PCs at the second stage.

List of Figures

Contents

Notations

	Dimensionality /Size related
d	:Pattern size or dimensionality
k	:Number of blocks or sub-patterns or sub-images of a pattern
k.r	:Dimensionality of locally-reduced pattern
m,n	:Number of rows and columns of an image pattern(matrix)
N	:Numebr of tarining patterns
r	:Locally-reduced sub-pattern size
u	:Block or sub-pattern size
w	:Globally-reduced patterns
	Pattern/Sub-pattern Data related
$(A_i)_{m\times n}$	:i^{th} Image pattern(matrix) of size $m \times n; i\epsilon 1, 2, ..., N$
$A = A_1, A_2, ..., A_N$	:Set of N training iamge patterns
$(A_i^j)_{m\times u}$	:j^{th} sub-image pattern of i^{th} image pattern,A_i
$(B_i)_{m\times kr}$	:Locally-reduced iamge pattern of$(A_i)_{m\times n}$
B=$B_1, B_2, ..., B_N$	:Set of locally-reduced image patterns of a
$(D_i)_{m\times w}$	:i^{th} Globally-reduced image pattern of $(B_i)_{m\times kr}$
D=$D_1, D_2, ..., D_N$	:Set of globally-reduced image patterns of B
$x_1, x_2, ..., x_d$	:d random feature variables of feature vector X
X	:Feature vector of d random feature variables $x_1, x_2, ..., x_d$
y	:transformed feature vector of X
$(X_i)_{d\times 1}$	:i^{th} Pattern;$i\epsilon 1, 2, ..., N$

X=$X_1, X_2, .., X_N$	:Set of N tarining patterns(or objects)
X^r=$X_1^r, X_2^r, ..., X_N^r$	:Set of N reduced training patterns by classical PCA
$(X_i^j)_{u\times 1}$	:j^{th} sub-pattern or Block of i^{th} pattern,$(X_i)_{d\times 1}$
$(Y_i)_{k.r\times 1}$	:i^{th} Locally-reduced pattern of $(X_i)_{d\times 1}$
$(Y_i^j)_{r\times 1}$	:Locally-reduced sub-pattern of $(X_i^j)_{u\times 1}$
Y=$Y_1, Y_2, ..., Y_N$	:Set of locally-reduced patterns of X
$(Z_i)_{w\times 1}$	:i^{th} Globally-reduced patterns of $(Y_i)_{k.r\times 1}$
Z=$Z_1, Z_2, ..., Z_N$	:Set of globally-reduced patterns of Y
$(C)_{d\times d}$	:Covariance matrix of patterns,X
$(C^g)_{(k.r)\times(k.r)}$	:(Inter-block)Covariance matrix of Y
$C_{q,t}^g$	:Matrix of inter-block covariances of R^q and R^t
$(c^j)_{(u\times u)}$or $(C^{ij})_{u\times u}$	:(Intra-block) sub-covariance matrix of j^{th} sub-patterns,P^j
$(M)_{n\times n}$	:Image covariance matrix of original iamges,A
$(E)_{d\times r}$	:Set of first r(¡d) global eigenvectors of C
$(E)_{n\times r}$	:Set of first r(¡n) global eigenvectors of M
G	:Set of all eigenvectors of C^g
V	:Combined matrix of local eigenvectors of all the blocks

Abbreviations

PCA	:Principal Component Analysis
2DPCA	:Two-Dimensional Principal Component Analysis
FP-PCA	:Feature Partitioning based PCA
FLPCA	:Flexible image Principal Component Analysis
IMPCA	:Image Principal Component Analysis
modPCA	:Modular PCA
MSE	:Mean Squared Error
PSNR	:Peak signal to noise ratio
PC	:Principal Component
SubPCA	:Sub-pattern based PCA
SSS	:Small Sample Size
SubXPCA	:Cross-Sub-pattern correlation based PCA
PV	:Projection Vector(Eigenvector selected for projection)
PR	:Pattern Recognition

Chapter 1

Introdiuction

In this chapter, we present a brief review of feature extraction(dimensionality reduction) and Principal Component Analysis.

1.1 Feature Extraction(Dimensionality Reduction)

Feature Extraction is an important stage of pattern recognition. Extraction of features is an important step and strongly influences classifier design. That is, if extracted features show significant differences from one class to another, the classifier can be designed more easily with better performance.Therefore, the extraction of features is a key issue in pattern recognition.

It is well known that often patterns contain a large number of features (e.g. a face image of 300×400 size consists of 12000 features). These patterns may contain features which are redundant or irrelevant to the classification task. Moreover, the pattern generating mechanism and the feature extraction techniques(transducers) are likely to introduce some distortion and noise, in addition to natural pattern variability. Therefore the fundamental task of feature extraction technoque is to extract most useful information from the original pattern and present it in a form of lower dimensionality vector, whose components represent the most significant features of input pattern.The goal of feature extraction technique is not merely dimensionality reduction, but to remove any redundant and irrelevant features

which may reduce the classifier performance. Further role of the feature extraction process is to establish whether it is necessary to seek additional features which would contain discriminatory information allowing the improvement of classifier performance.

Most feature extraction techniques compress the observed information into a lower dimension space to facilitate its transmission or storage or classification. Here the elimination of the irrelevant information and redundancy(using feature extraction techniques) is an integral part of the transformation which maps original pattern vector $\mathbf{X}$ into $\mathbf{Y}$, the tranformed vector, a new lower dimensional feature space which is given by

$$Y = f(X)$$

The mapping f(...) is obtained by optimizing a criterion function J(...).

As the number of dimensions increases the generalization ability is likely to come down for finite training data. Feature extraction may be seen as the one which improves generalization ability of recognition system by keeping number of dimensions minimum.

Some commonly used feature extraction techniques include Principal Component Analysis(PCA), Linear Discriminnat Analysis, Independent Componet Analysis, Projection Pursuit, Random Projections.

Here we shall focus on PCA henceforth in this work.

1.2 Classical Principal Component Analysis(PCA)

Principal Component Analysis(PCA) is one of the widely used techniques for dimensionality reduction with widespread applications to data reduction, image processing, visualization, pattern recognition, exploratory data analysis, etc.

PCA concerned with explaining the *variance-covariance structure* through a few linear combinations of the original features. Although d components are required to reproduce the total system variability, often much of this variability can be accounted for by a small number, $r(< d)$ of the Principal Components (PCs) from the original d features. The N Measurements with r PCs are used to replace original N Measurements with d features.

1.2.1 How to Perform PCA on Given Set of Patterns ?

Consider $X = X_1, X_2, ..., X_N$, the set of N training patterns of dimensionality d.
The steps of classical PCA are given as follows

1. Perform a *mean subtraction* operation upon the training data patterns

2. Calculate the Covariance Matrix.

$$(C)_{d\times d} = \frac{1}{N}.\sum_{i=1}^{N}[X_i - \bar{X}].[X_i - \bar{X}]^T \tag{1.1}$$

   Where $\bar{X}$ is the mean of tarining patterns , X.

3. Determine eigenvalues, λ and eigenvectors, **e** of the covariance matrix, **C** such that

$$(C)_{d\times d}.(e)_{d\times 1} = (e)_{d\times 1}.(\lambda)_{1\times 1} \tag{1.2}$$

   If the rank of **C** is **d** then one can find **d** eigenvalues.

4. Sort the eigenvalues and the corresponding eigenvectors in non-increasing order.

5. Choose first $r(< d)$ column eigenvectors, (denoted by $E_{d\times r}$) and project the data set **X** on $E_{d\times r}$ to get the reduced data X^r .

$$X^r_{N\times r} = X_{n\times d}.E_{d\times r} \tag{1.3}$$

   The original data can be reconstructed with most of the variance(information) from compressed or reduced data $\widehat{X}_{N\times d}$ is given by

$$X^r_{N\times r} = X^r_{N\times r}.E^T_{r\times d} \tag{1.4}$$

   Where $\widehat{X}_{N\times d}$ is reconstructed data.

1.2.2 Why is PCA so Popular in Dimensionality Reduction ?

It is observed that the use of PCA is widespread for dimensionality reduction.
Here we enumerate some reasons.

- PCA is a global scheme because it considers covariances (correlations) between every pair of original d variables (features). Therefore it is highly effective in dimensionality reduction when global variations are prominent.
- PCA is an optimal linear scheme (in terms of maen squared error) for compressing a set of high dimension vectors into a set of low dimension vectors and for reconstructing the original vectors
- The model parameters (eigenvectors and eigenvalues) can be computed directly from the data i.e. by diagonalizing the sample covariance.
- Compression and Decompression are easy operations and they require only matrix multiplications.
- Further, PCA minimizes representation entropy and it makes output variables (PCs) mutually uncorrelated.

1.2.3 Fundamental Problems /Issues with Classical PCA

Although PCA is popular in its application, there are several issues which inhibit an analyst. Here we describe a few of them based upon our study.

1. *High computational complexity.* The classical PCA methods are not suitable for high dimensional data (d) or large number of data points (N), because PCA needs high computational requirements for such data in particular for huge image data. Computing the sample covariance matrix or correlation matrix itself is expressive ($O(N.d^2)$).
2. *Poor performance(that is less generalization ability) with data of prominent local variations.* PCA is global feature extraction technique which may perform well when global variations among patterns are dominant, however it may not perform well when local variations among patterns are dominant.

3. *Small sample size(SSS) problem.* PCA may not be good to reduce dimensions in the case of a Small Smaple Size (SSS) problem. That is when number of training samples are less as compared to number of features or dimensions, PCA may not effectively perform dimensionality reduction or feature extraction

4. *Not good for non-linear data.* Further, Classical PCA is suitable in applications where the underlying structure is linear. The linear PCA either needs more principal components or unsuitable for the data sets where nonlinear structure is present.

5. *choosing right number of principal components.* Right choice of principal components influence classifier performance as well as the total amount of variance (structure) in the reduced data. Now the question arises as to how to choose right number of Principal components ?

In the next chapter, we discuss the other PCA methods to address the problems faced by classical PCA.

Chapter 2

Principal Component Analysis Methods: A Literature Survey

2.1 Introduction

In this chapter we review the literature related to Principal Componet Analysis (PCA) methods in brief. For better understanding we classify the literature into catagories like, Feature Partitioning or Block based PCA (FP-PCA) methods, 2D structure based PCA methods.

2.2 2DPCA

2DPCA[6] is based on 2D image matrices where as PCA is based on 1D vectors, so the image matrix does not need to be transformed into a vector prior to featue extraction. Instead, an *image covariance matrix* is constructed directly using the original image matrices, and its eigenvectors are derived for image feature extraction.

Algorithm

1. Let X denote an n-dimensional unitary column vector. Our idea is to project image A, an $m \times n$ random matrix, onto X by the following linear transformation:

$$Y = AX. \tag{2.1}$$

Thus, we obtain an m-dimensional projected vector Y, which is called the projected feature vector of image A. How do we determine a good projection vector X? In fact, the total scatter of the projected samples can be introduced to measure the discriminatory power of the projection vector X. The total scatter of the projected samples can be characterized by the trace of the covariance matrix of the projected feature vectors. From this point of view, we adopt the following criterion:

$$J(X) =_{tr} (C), \tag{2.2}$$

where **C** denotes the covariance matrix of the projected feature vectors of the training samples and $_{tr}(C)$ denotes the trace of **C**. The physical significance of maximizing the criterion in 2.2 is to find a projection direction X, onto which all samples are projected, so that the total scatter of the resulting projected samples is maximized.

2. Compute covarince matrix: The covarince matrix **C** can be denoted by,

$$C = E(Y - EY)(Y - EY)^T \tag{2.3}$$

$$= E[AX - E(AX)][AX - E(AX)]^T \tag{2.4}$$

$$= E[(A - EA)X][(A - EA)X]^T \tag{2.5}$$

so,

$$_{tr}(C) = X^T[E(A - EA)^T(A - EA)]X \tag{2.6}$$

$$G_t = E[(A - EA)^T(A - EA)] \tag{2.7}$$

Suppose that there are M training image samples in total, the j^{th} training image is denoted by an $m \times n$ matrix $A_j(j = 1, 2, ..., M)$, and the averageimage of all training

samples is denoted by $\bar{A}$. Then ,G_t can be evaluated by

$$G_t = \frac{1}{M}\sum_{j=1}^{M}(A_j - \bar{A})^T(A_j - \bar{A}). \tag{2.8}$$

so,the criterion is

$$J(X) = X^T G_t X \tag{2.9}$$

the optimal projection axis X_{opt} is the unitary vector that maximizes J(X), i.e., the eigenvector of G_t corresponding to the largest eigenvalue.

3. Feature extraction: The optimal projection vectors of 2DPCA, $E_k = (e_1, e_2, ...e_d)$, are used for feature extraction.For a given image sample A, let

$$Y_k = AE_k, k = 1, 2, ..., d \tag{2.10}$$

Then,we obtain a family of projected feature vectors, $Y_1, ..., Y_d$, which are called the *Principal Componet(vectors)* of the sample image A. // The principal component vectors obtained are used to form an $m \times d$ matrix $B = [Y_1, ..., Y_d]$, which is called the *feature matrix* or *featue image* of the image sample A. This feature image is used for classification process

2.2.1 Comparison and Properties of 2DPCA

2DPCA[6] has many advantages over conventional PCA (Eigenfaces). In the first place, since 2DPCA is based on the image matrix, it is simpler and more straightforward to use for image feature extraction. Second, 2DPCA is better than PCA in terms of recognition accuracy in all experiments. Although this trend seems to be consistent for different databases and conditions, in some experiments the differences in performance were not statistically significant. Third, 2DPCA is computationally more efficient thanPCA and it can improve the speed of image feature extraction significantly. However, it should be pointed out that 2DPCA-based image representation was not as efficient as PCA in terms of storage requirements, since 2DPCA requires more coefficients for image representation than PCA.

2.3 Feature Partitioning or Block based PCA (FP-PCA) Methods

It is now known that classical PCA suffers from the drawbacks of not coping well with high dimensional data scaling up to large data set due to its prohibitive computational complexity ($O(N.d^2)$). Another shortcoming is that classical PCA may not perform well in terms of recognition for applications where local region based features have discriminant information (e.g.facial expressions, pose, illuminations, etc and change detection applications). To overcome these problems, Block-based PCA methods were emerged.

2.3.1 modPCA

In modPCA[5] technique, the face images are divided into smaller sub-images and the PCA approach is applied to each of these sub-images.

Algorithm

1. Each image in the training set is divided into N smaller images. Hence the size of each sub-image will be $\frac{L^2}{N}$.

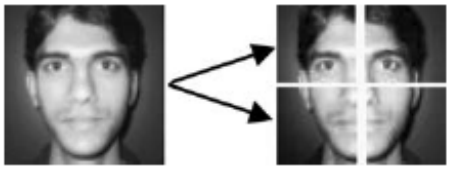

Figure 2.1: A Face image divide into N smaller images, where N=4

These sub images can be mathematically represented as

$$I_{ij}(m,n)i = I_i(\frac{L}{\sqrt{N}}(j-1)+m, \frac{L}{\sqrt{N}}(j-1)+n) \ \forall i,j \tag{2.11}$$

where i varies from 1 to M, M being the number of images in the training set, j varies from 1 to N, N being the number of sub-images and m and n vary from 1 to $\frac{L}{\sqrt{N}}$

2. The average image of all the training sub-images is computed as

$$A = \frac{1}{M.N} \sum_{i=1}^{M} \sum_{j=1}^{N} I_{ij} \tag{2.12}$$

3. Normalize each training sub-image by subtracting it from the mean as

$$Y_{ij} = I_{ij} - A \forall i, j \tag{2.13}$$

4. From the normalized sub-images the covariance matrix is computed as

$$C = \frac{1}{M.N} \sum_{i=1}^{M} \sum_{j=1}^{N} Y_{ij}.Y_{ij}^{T} \tag{2.14}$$

5. Next we find the eigenvectors of C that are asso- ciated with the k largest eigenvalues. We repre- sent the eigenvectors as $e_1, e_2, ..., e_u$. The weights are computed from the eigenvectors as shown below

$$W_{pnjK} = E_K^T (I_{Pnj} - A) \forall p, n, j, K \tag{2.15}$$

6. Where K takes the values 1,2,...,u, n varies from 1 to Γ, Γ being the number of images per individual, and p varies from 1 to P, P being the number of individuals in the training set. Weights are also computed for the test sub-images using the eigenvectors as shown in the next equation:

$$W_{test.jK} = E_K^T (I_{test.jK} - A) \ \forall j, K \tag{2.16}$$

7. Mean weight set of each class in the training set is computed from the weight sets of the class as shown below:

$$T_{pjK} = \frac{1}{\Gamma} \sum_{K=1}^{u} \sum_{n=1}^{\Gamma} W_{pnjK} \forall p, j \tag{2.17}$$

8. Next the minimum distance is computed as shown below:

$$D_{pj} = \frac{1}{u} \sum_{K=1}^{u} |W_{test.jK} - T_{pjK}| \tag{2.18}$$

$$D_p = \frac{1}{N} \sum_{j=1}^{N} D_{p_j} \tag{2.19}$$

$\min(D_p < \Theta_i)$ for a particular value of p, the corresponding face class in the training set is the closest one to the test image. Hence the test image is recognized as belonging to the pth face class.

Comparison and properties of modPCA

A modular PCA method, which is an extension of the PCA method for face recognition. The modular PCA method performs better than the PCA method under the conditions of large variations in expression and illumination.For large variations in pose there is no significant improvement in the performance of modular PCA. For face recognition, the modular PCA method can be used as an alternative to the PCA method. In particular, the modular PCA method will be useful for identification systems subjected to large on Computer Vision and Pattern Recognition.

2.3.2 SubPCA

The sub-pattern based PCA (SubPCA[9]) technique which divides each pattern into equally-sized sub-patterns and groups sililar sub-patterns from all patterns into corresponding sub-pattern set. Local features are extracted from each sub-pattern set and are concatenated to form reduced patterns.

Algorithm

Partition the given mean-corrected data$(X)_{n\times d}$ and apply classical PCA

1. Partition datum, X_i into ($k \geq 2$) equally-sized sub-patterns. Each sub-pattern of size d_l, Where $d_l = \lfloor \frac{d}{k} \rfloor$, SP_j is the set of j^{th} sub-patterns of X_i, $\forall i = 1, .., N$.

2. For every sub-pattern set, SP_j , Where j=1,..,k, repeat the following steps
   (i)Compute covariance matrix, $(C_j)_{d_l\times d_l}$.
   (ii)Compute eigenvalues (λ_p^j) and corresponding eigenvectors (e_p^j), where p=1,..,d_l.
   (iii)Select $r(< d_l)$ eigenvectors corresponding to the first r largest eigenvalues obtained in step(ii). Let E_j be the set of r eigenvectors(column vectors) selected in this step.
   (iv)Extract r local features (PCs) from SP_j by projecting SP_j onto $(E_j)_{d_l\times r}$ as follows. Let Y_j be the reduced data in this step and is given as

$$(Y_j)_{N\times r} = (SP_j)_{N\times d_l}(E_j)_{d_l\times r}$$

3. Concatenate $Y_j, \forall j = 1, .., k$, as shown below. Let Z denote such combined data.Z_i is the i^{th} row of $(Z)_{N\times kr}$, which corresponds to X_i and is given as

$Z_i = [y_1(i,1), y_1(i,2), ..., y_1(i,r)$,...and so on ..,
$y_k(i,1), ..., y_k(i,r)]$ and $(y_j(i,1), y_j(i,2), ..., y_j(i,r))$ is the i^{th} row of Y_j.
Here Z is given for the classification.

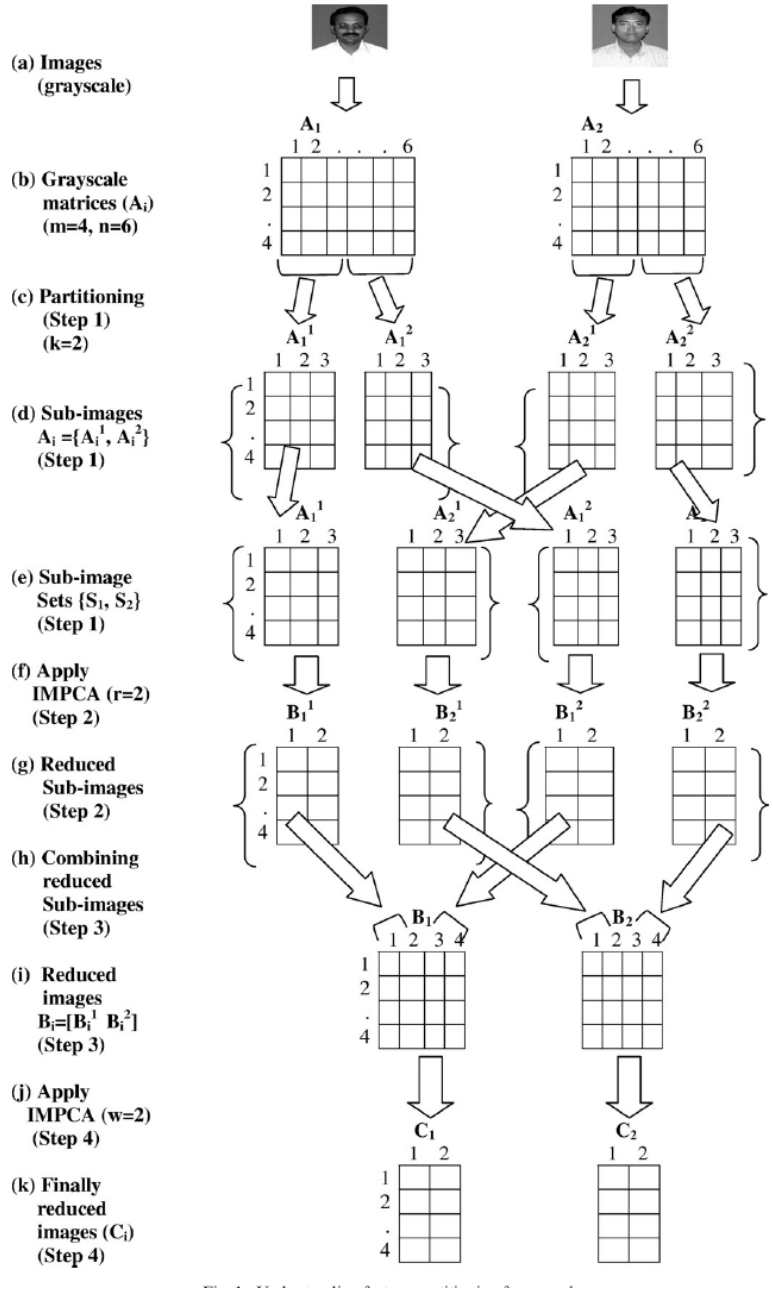

Figure 2.2: SubPCA Algorithm

Comparison and properties of SubPCA

SubPCA does not consider correlations across different sub-patterns which may result in poor summarization of variance (and hence low dimensionality reduction) in comparison to PCA(since PCA considers entire correlation structure of the data). Noisy and less expressive features contain less variance. Such noisy features may reduce classification rate as well. In SubPCA, there is possibility of relatively more(possibly correlated) noisy or less-expressive features which are spread across different sub-patterns, and there is no mechanism to retain them. Thus classification accuracies of SubPCA may not be encouraging, then global variations are predominant in the data. Since PCA exploits the entire correlation structure in the data(hence it is global scheme), it shows better summariztion of variance (which implies high dimensionality reduction) and may lead to better classifiaction rates in comparison to SubPCA, when global variations and prominent. Interestingly, SubPCA performs better than PCA in terms of classification when the local structure (local features) has good discrimination information. Another advantage of SubPCA is, it shows much lesser time complexity as compared to PCA.

2.3.3 SubXPCA

In this section, we take a vital issue *loss of inter-block correlations*, which is crucial for dimensionality reduction and classification. SubXPCA[10] overcomes the problems faced by global and local feature extraction methods to PCA, keeping the merits of both intact. SubXPCA does so, (i) by extracting local features from sub-patterns(blocks) (local feature extraction) and (ii) by exploiting inter-sub-pattern correlations (cross-sub-pattern correlations) among those locally extracted features (global extraction). SubXPCA balances the global PC computation of classical PCA against the local viewpoint of SubPCA and similar methods.

SubXPCA Algorithm

1. *Partitioning step*(Step-1 in Fig 3.1):

   Divide every d-dimensional pattern $X_i, \forall i = 1, 2, ..., N$ into $k(\geq 2)$ equally-sized sub-patterns $X_i^1, X_i^2, ..., X_i^k$. Each sub-pattern is of size u, where $u = \lfloor \frac{d}{k} \rfloor$.we consider equally-sized sub-pattern option for simplicity.Sub-pattern size is chosen to minimize the loss of last features. However ,it is not mandatory to consider equally-sized sub-patterns.Sub-patterns are formed by choosing features contiguously as they appear in the pattern. In other words, for a given pattern,$X_i = (x_{i1}, x_{i2}, ..., x_{id})^T$, first sub-pattern X_i^1 contains features, $x_{i1}, x_{i2}, ..., x_{iu}$ and j^{th} sub-pattern, X_i^j contains features given by

   $$(X_i^j)_{u\times 1} = (x_{il}, x_{i(l+1)}, ..., x_{i(l+u1)})^T \tag{2.20}$$

   Where $l = (j-1).u + 1, 1 \leq i \leq N, 1 \leq j \leq k$

2. *Grouping step*(Step -2 in fif 3.1):

   We pick-up j^{th} sub-pattern, corresponding to every pattern, X_i;i=1,2,...,N, and form j^{th} sub-pattern set (or j^{th} sub-pattern group), P^j is given by

   $$(P^j)_{N\times u} = [X_1^j X_2^j ... X_n^j]^T \tag{2.21}$$

   Here we use the option of grouping of homogeneous sub-patterns

3. *Local feature extraction step*(*step-3 in Fig. 3.1*).

   For every sub-pattern set P^j , where j=1,2,...,k, repeat the following steps(a)-(d)

   (a) Compute local variance matrix, $(\mathbf{C}^j)_{u\times u}$ as given by

   $$(C^j)_{u\times u} = \frac{1}{N}.\sum_{i=1}^{N}[X_i^j]_{u\times 1}.[X_i^j]_{1\times u}^T \tag{2.22}$$

   (b) Compute eigenvalues (λ_p^j) and corresponding eigenvectors (e_p^j) . where p=1,2,...,u, using eigen value decomposition (EVD) of C^j given by

   $$C^j.e_p^j = e_p^j.\lambda_p^j \tag{2.23}$$

   (c) Select $r(\leq u)$ eigenvectors corresponding to the first r largest eigenvalues obtained in the preceding step. Let E^J be the set of r eigenvectors (column vectors) selected

in this step and is given as follows

$$(E^j)_{u\times r} = [e_1^j e_2^j ... e_r^j]_{u\times r} \tag{2.24}$$

(d) Extract r local features (local PCs) by projecting P^j onto E^j as follows. Let R^J be the reduced data in this step and is given as follows

$$(R^j)_{N\times r} = (P^j)_{N\times u}.(E^j)_{u\times r} \tag{2.25}$$

$$(R^j)_{N\times r} = \begin{bmatrix} (Y_1^j)^T \\ (Y_2^j)^T \\ . \\ . \\ . \\ (Y_3^j)^T \end{bmatrix} = \begin{bmatrix} y_1^j(1) & y_1^j(2) & ... & y_1^j(r) \\ y_2^j(1) & y_2^j(2) & ... & y_2^j(r) \\ . & . & ... & . \\ . & . & ... & . \\ . & . & ... & . \\ y_N^j(1) & y_N^j(2) & ... & y_N^j(r) \end{bmatrix} \tag{2.26}$$

where Y_i^j is the locally-reduced version of X_i^j , the j^{th} sub-pattern of X_i

4. *Combining locally-extracted features step* (Step-4 in Fig 3.2):

(a) Form locally-reduced pattern, Y_i by concatenating locally-reduced sub-patterns (local features), $(Y_i^j)_{r\times 1}, \forall j = 1, 2, ..., k$, as shown below.

$$(Y_i)_{k.r\times 1} = [(Y_i^1)^T, (Y_i^2)T, ..., (Y_i^K)T]_{k.r\times 1}^T \tag{2.27}$$

(b) Perform global feature extraction using cross(inter)-sub-pattern correlations(covariances) of $Y = Y_1, Y_2, ..., Y_N$, obtained in the preceding step as given in (i)-(iv).

(i) Compute global covariance matrix , $(C^g)_{(k.r)\times(k.r)}$ for the data Y as follows

$$(C^g)_{(k.r)\times(k.r)} = \frac{1}{N}.\sum_{i=1}^{N}[Y_i].[Y_i]^T \tag{2.28}$$

(ii) Compute eigenvalues (λ_s) and corresponding eigenvectors (e_S) , where s=1,2,...(k.r) using eigenvalue decomposition given by

$$C^g.e_s = e_s.\lambda_s \tag{2.29}$$

(iii) Select $w(\leq k.r)$ eigenvectors corresponding to first w largest eigenvalues obtained in the preceding step. Let E^g be the set of w eigenvectors selected in this step and is given by

$$(E^g)_{k.r\times w} = [e_1 e_2 ... e_w]_{k.r\times w} \tag{2.30}$$

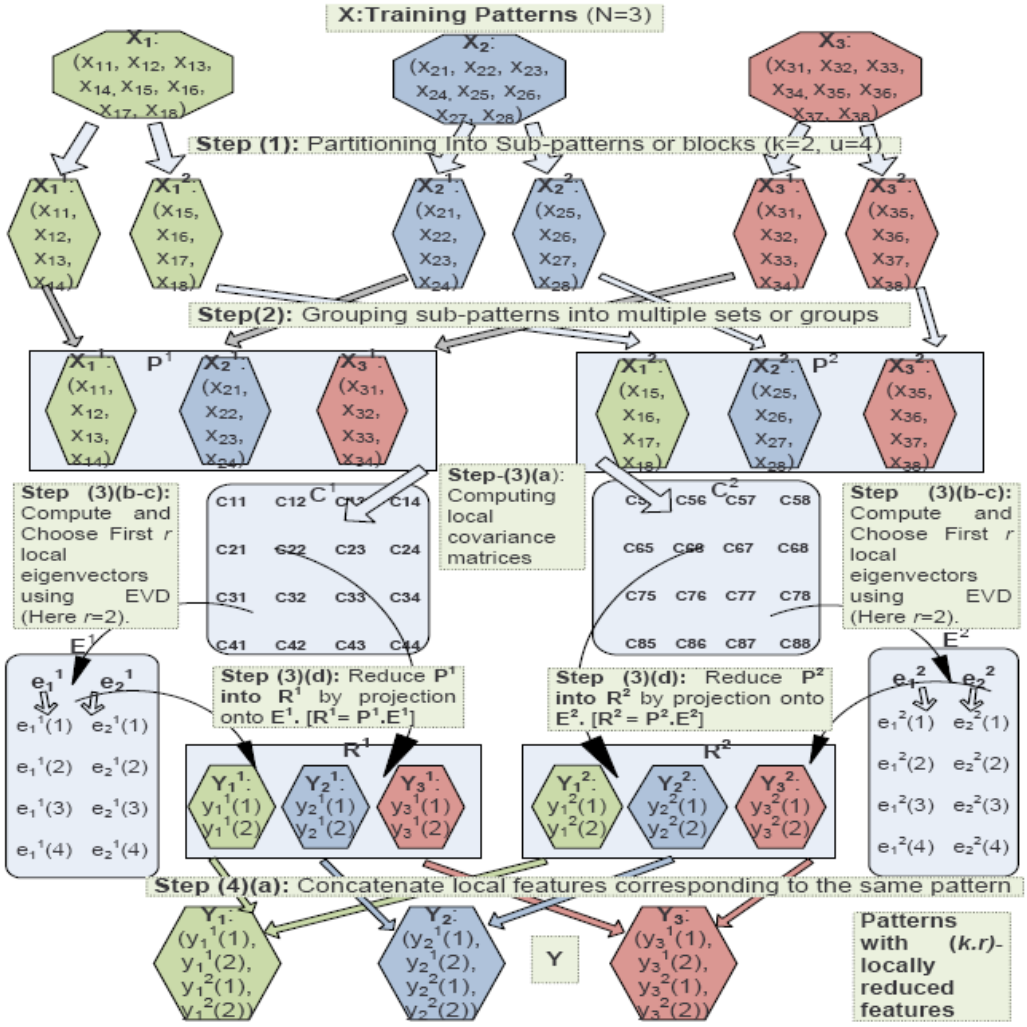

Figure 2.3: SubXPCA Part-1

(iv) Extract w global features (global PCs) by projecting Y (obtained in step-4(a)) onto E^g . Let Z be the data obtained after projection in this step and is given as

$$(Z)_{N\times w} = (Y)_{N\times k.r}.(E^g_{k.r\times w}) \tag{2.31}$$

We finally obtained $(Z)_{N\times w}$ which is the reduced form of $(X)_{N\times d}$ and Z is further used for subsequent tasks such as classification, recognition, clustering, etc.

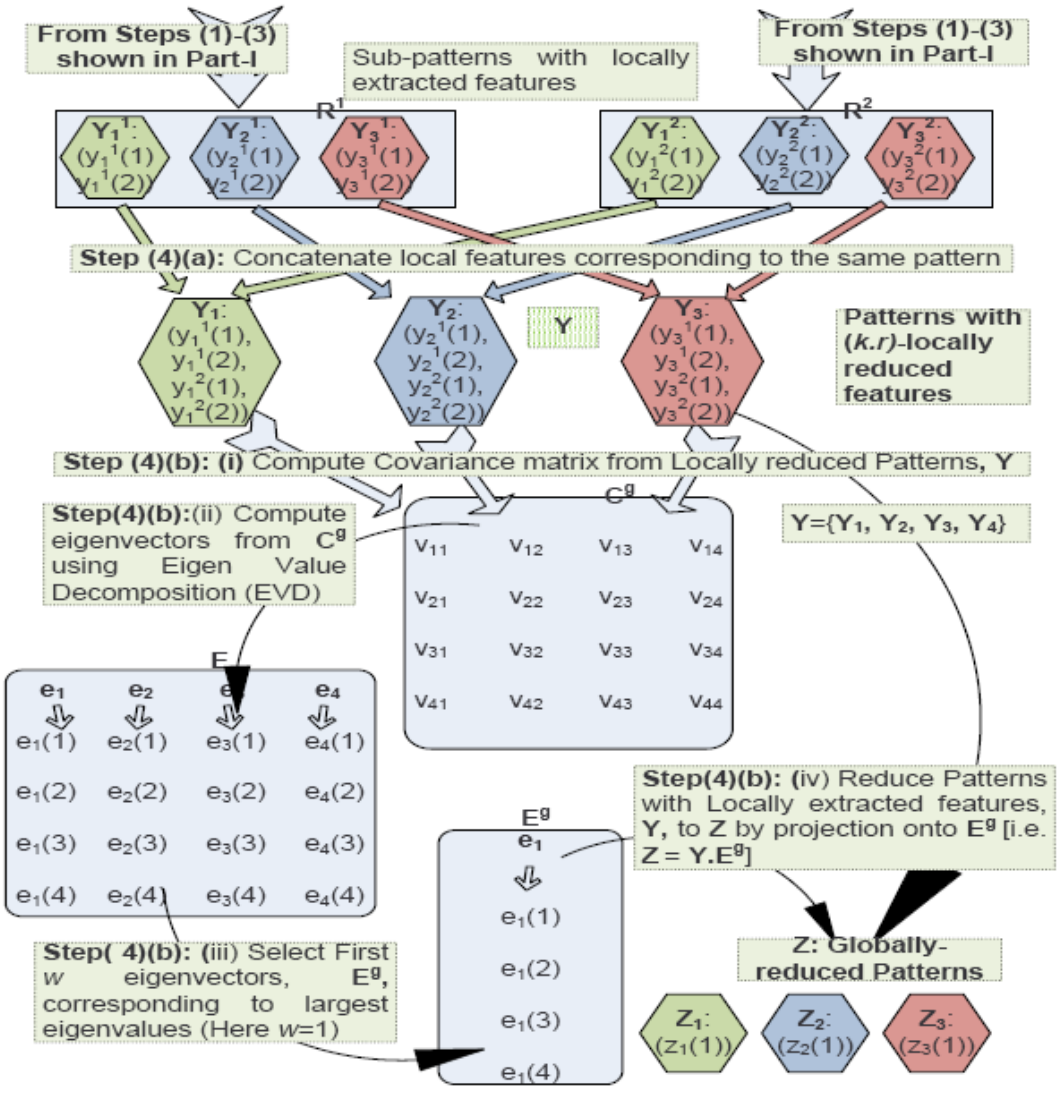

Figure 2.4: SubXPCA Part-2

2.4 Summary

In This Chapter we briefly explained the Different types of PCA methods. In this chapter, the proposed FP-PCA approach, SubXPCA technique as a solution to the issues 'loss of inter-sub-pattern correlations' and 'feature order dependency'. SubXPCA improves the summarization of variance, reduces the dimensionality and improves the classification accuracy, thus saving computational and processing costs. In the next chapter we discuss about the Comaprison of PCA approaches for face image dimension.

Chapter 3

Comparison of FP-PCA approaches for Face Image

3.1 A relative comparison between SubXPCA approach and Others

In this section, the relative performance of SubXPCA approach is analysed.

3.1.1 SubXPCA Versus SubPCA

PCA is able to retain meaningful information (variance) in the major axes (eigenvectors corresponding to largest eigenvalues), where as variance associated to experimental error, measurement inaccuracy, and/or rounding (i.e.noise) is summarized in minor axes (eigenvectors corresponding to smaller eigenvalues). SubPCA considers all $k.r$ local features (extracting r(¡u) from each of k sub-patterns), which may also include less-expressive, noisy or correlated features. SubPCA ignores correlations among these local features which may result in low dimensionality reduction due to poor summarization of variance. SubXPCA exploits these correlations to summarize the variance among those $k.r$ local features and select a few salient features, $w(< k.r)$, thus results in better summarization of variance and high dimensionality reduction. Those w features may increase the classification accuracy as well.

SubXPCA is relatively more feature order independent as compared to SubPCA and SubX-PCA shows better performance as compared to SubPCA with overlapping sub-patterns. An interesting characteristic of SubXPCA is that, it is able to do well, when global variations are prominent, in which case SubPCA does not perform well.

3.1.2 SubXPCA Versus PCA

As in the Step-3 of SubXPCA , PCA can be derived as a special case of subXPCA, where r, u are eigenvectors chosen from each sub-pattern set and sub-pattern size respectively. This as the number of local eigenvectors(r) in each sub-pattern increases, SubXPCA very closely approaches PCA in terms of summarization of variance. An interesting characteristic of SubXPCA is that, SubXPCA is able to show better performance when local variations are prominent, in which case PCA does not perform well.

3.2 Discussion : About the Performance of SubXPCA method with eigenvectors corresponding to lowest and random eigenvalues

In SubXPCA Algorithm, described in section 2.3.3, a question was raised. ***Can the features found in the stage*** 1^{st} ***affect the overall performance of the SubXPCA approach?***

Accordingly we made two experiments where we may choose eigenvectors (a) at random or (b) from those associated with small eigenvalues (minor components), while the second stage is meant to ensure overall PCA property at global level. Here our approach might remove detail at local level (i.e. first level). Select $r(\leq u)$ eigenvectors corresponding to the first r lowest eigenvalues or r random eigenvalues obtained in the step-3-b of SubXPCA algorithm. Let E^J be the set of r eigenvectors (column vectors) selected in this step and is given as follows.

$$(E^j)_{u\times r} = [e_1^j e_2^j ... e_r^j]_{u\times r} \tag{3.1}$$

Therefore here we have these new configuration of SubXPCA experimentally by two approaches.

(a) By testing classification accuracy using nearest neighbor approach

(b) By observing differences in PSNR value

These approaches are tested and reported in the following. For the experiments, we consider eight subjects each has eight images as training data and one subject which has six images as testing data from Yale Database.

3.2.1 Eigenvectors corresponding to lowest eigenvalues

The following chart shows classification accuracy for SubXPCA method for eigenvectors associated to small eigenvalues. The accuracy is very less if we take eigenvectors corresponding to small eigenvalues.

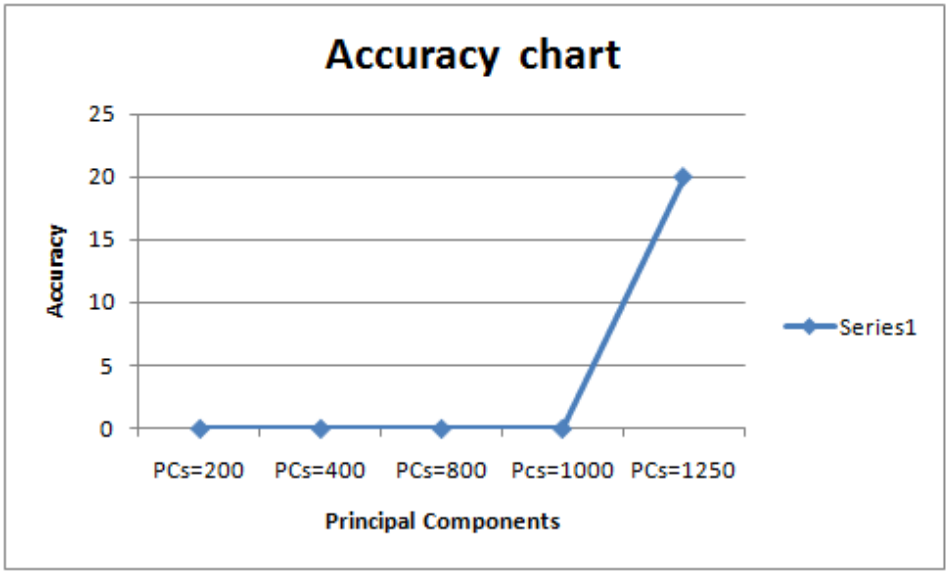

Figure 3.1: Accuracy graph when chosen eigenvectors associated with small eigenvalues

3.2.2 Eigenvectors corresponding to random eigenvalues

The following chart shows classification accuracy for SubXPCA method for eigenvectors associated to random eigenvalues. We can not assure good accuracy rating for random eigenvectors.

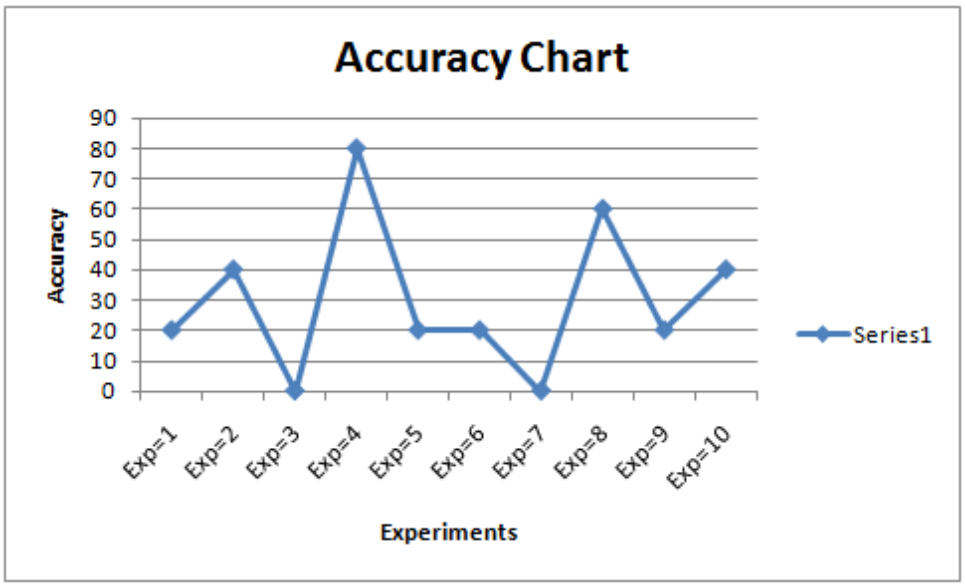

Figure 3.2: Accuracy graph when chosen eigenvectors associated with random eigenvalues

3.3 Summary

We presented brief comparison of FP based PCA approaches, and experimented SubXPCA method with small and random eigenvalues.

Chapter 4

Experiments and Results

In this chapter we discuss about how to reconstruct a image, what is PSNR and information preservation by each PCA method through PSNR values.

4.1 Reconstruction of a Image

Suppose the orthonormal eigenvectors corresponding to the first d largest eigenvectors of the image covariance matrix G_t are $X_1, ..., X_d$. After the image samples are projected onto these axes, the resulting principal components vectors are $Y_k = AX_k(k = 1, 2, ..., d)$. Let $V = [Y_1, ..., Y_d]$ and $U = [X_1, ..., X_d]$ then

$$V = AU \tag{4.1}$$

Since $X_1, ..., X_d$ are orthonormal, from (4.1), it is easy to obtain the reconstructed image of sample A:

$$\tilde{A} = VU^T = \sum_{k=1}^{d} Y_K X_k^T \tag{4.2}$$

Let $\tilde{A}_k = Y_k X_k^T (k = 1, 2, ..., d)$, which is of the same size as image A, and represents the *reconstructed image of A*.

4.2 Peak Signal to Noise Ratio(PSNR)

4.2.1 What is PSNR

The PSNR [1]is most commonly used as a measure of quality of reconstruction of Image. The signal in this case is the original Image, and the noise is the error introduced by reconstruction. When comparing reconstruction it is used as an approximation to human perception of reconstruction quality, therefore in some cases one reconstruction may appear to be closer to the original than another, even though it has a lower PSNR (a higher PSNR would normally indicate that the reconstruction is of higher quality).

It is most easily defined via the mean squared error (MSE) which for two $m \times n$ monochrome images I and K where one of the images is considered a noisy approximation of the other is defined as:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2 \tag{4.3}$$

The PSNR is defined as

$$PSNR = 10.log_{10}(\frac{MAX_I^2}{MSE}) \tag{4.4}$$

$$PSNR = 20.log_{10}(\frac{MAX_I^2}{MSE}) \tag{4.5}$$

Here, MAX_I is the maximum possible pixel value of the image. If it is gray image the pixel value is 255. PSNR is computed for original image against the reconstructed image. Here we use PCA methods ClassicalPCA, SubXPCA, SubPCa, modPCA and 2DPCA to reconstruct a image. The following are graphs drawn for PSNR values against the number of PCs for all PCA methods applied for five different subjects each has five different images from Yale database.

4.3 Results

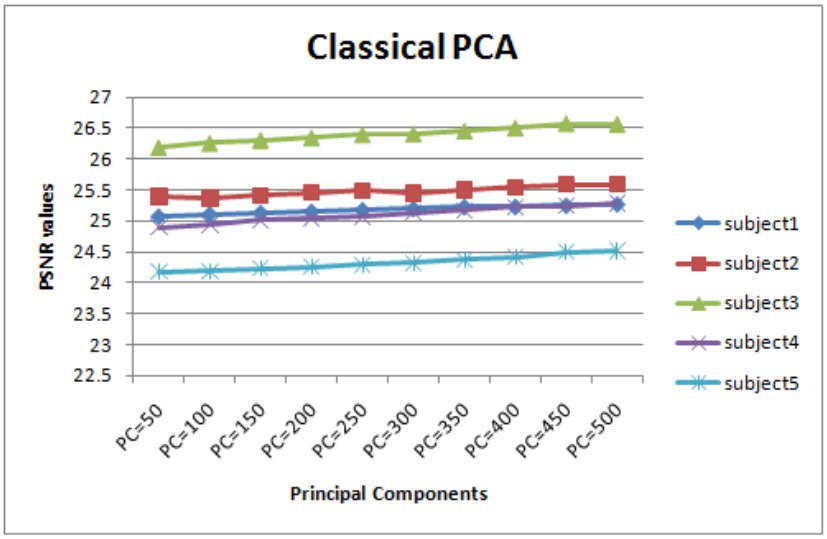

Figure 4.1: PSNR values of classical PCA for five different subjects from yale database

From the above graph it is observed that the PSNR values increase with increase in number of Principal Components.

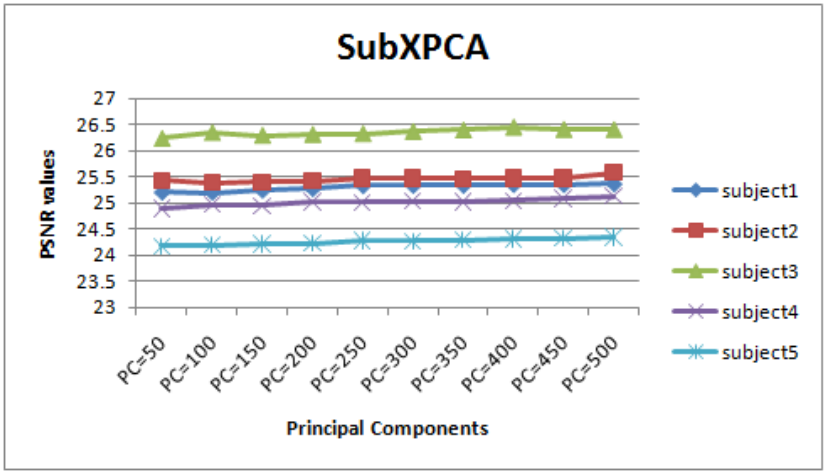

Figure 4.2: PSNR values of SubXPCA for five different subjects from yale database

From the above graph it is observed that the PSNR values increase with increase in number of Principal Components.

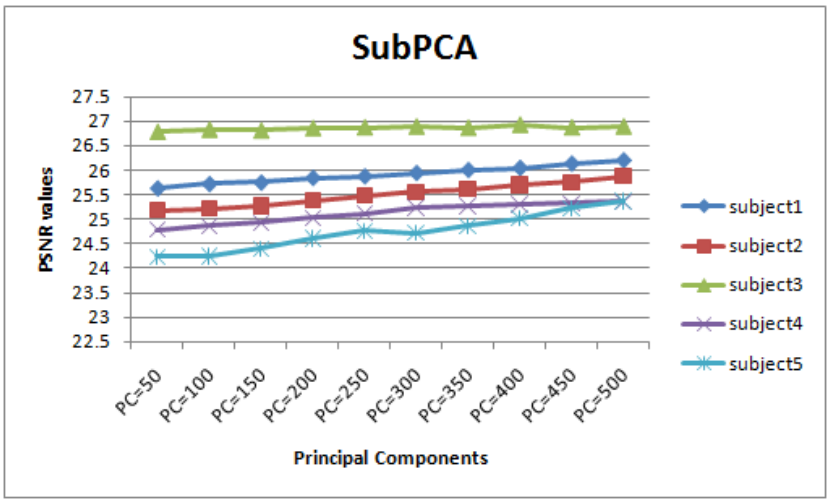

Figure 4.3: PSNR values of Sub PCA for five different subjects from yale database

From the above graph it is observed that the PSNR values increase with increase in number of Principal Components.

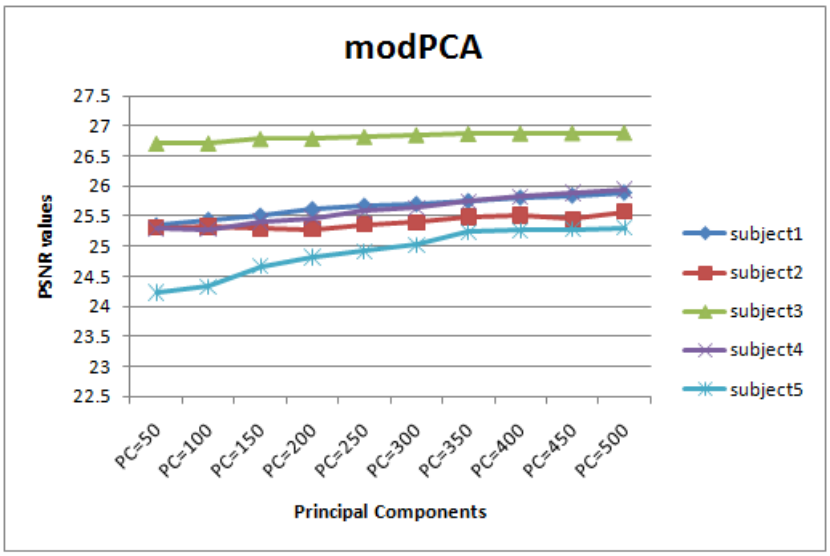

Figure 4.4: PSNR values of ModPCA for five different subjects from yale database

From the above graph it is observed that the PSNR values increase with increase in number of Principal Components.

From the above graph it is observed that the increasing rate of PSNR values is not

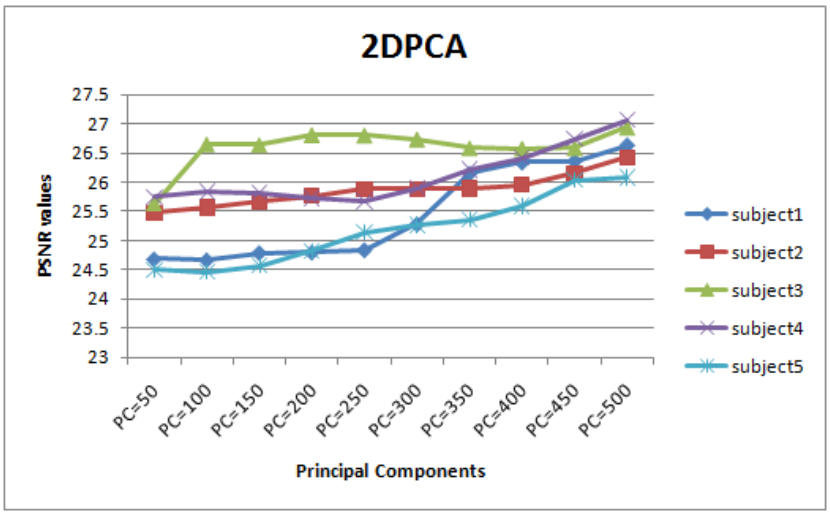

Figure 4.5: PSNR values of 2DPCA for five different subjects from yale database

constant.

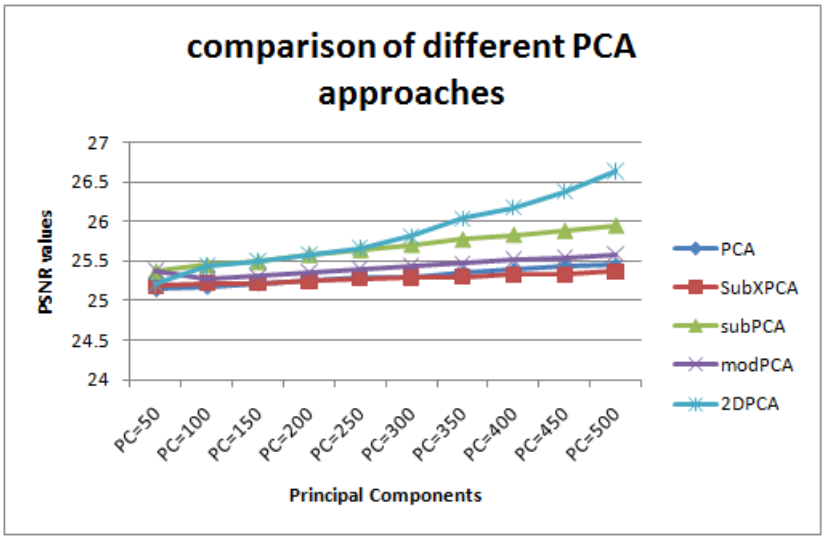

Figure 4.6: PSNR values of PCA methods

From the above graph it is observed that 2DPCA gives highest PSNR value, when compared to with other PCA approaches.

From the above graph, it is observed that the computation time for classicalPCA is more and for 2DPCA it is less, compared to all other methods.

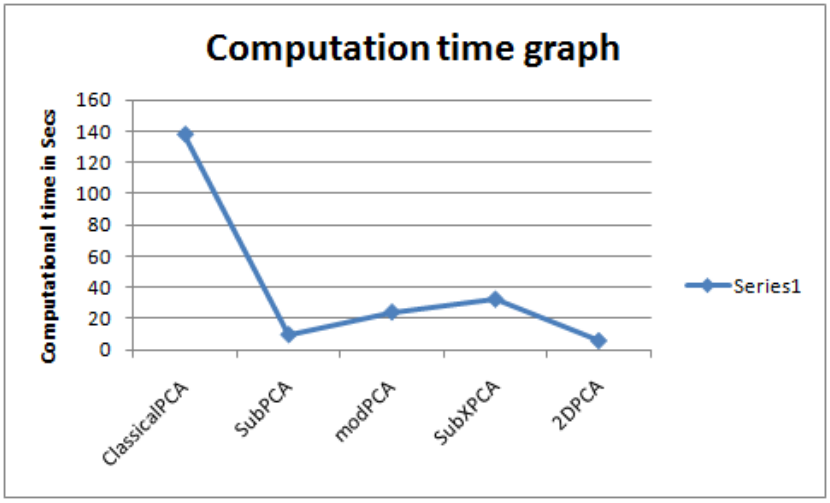

Figure 4.7: Computation time Graph for reconstruction by PCA methods

Chapter 5

Conclusions

From the experimentation the following observations are seen.

- **Relative highest PSNR values**: 2DPCA gives highest PSNR value, when compared to with other PCA approaches. The relative PSNR value changes from image to image and method to method.
- **What happens if more PCs are added**: The PSNR value is increased when more number of PCs are added. The PSNR value is also depends on the quality of original image.
- **How many PCs are needed for good reconstruction**: The image quality is seen to be increased in number of PCs. We can increase PC number upto the size of image, however with a loss in addition computational time that is neede and the loss of benefit of dimensionality reduction.

- **PSNR Vs Visual Inspection**:

Figure 5.1: PSNR Vs Visual inspection

The above figure showing that the PSNR value is high for a image which is visullay not good,So the case in which image is visually good the PSNR value might be low.

Bibliography

[1] http://en.wikipedia.org/wiki/Peak_signal-to-noise_ratio.

[2] A.Negi and K.Vijaykumar. SubXPCA versus PCA : A theoretical investigation. *Procs.of International Conference on Pattern Recognition,Turkey*, pages 4170–4173, 2010.

[3] Sabine Susstrunk Clement Fredembach, Michael Schroder. Eigenregions for image classification. *IEEE Transactions on Pattern analysis and Machine Intelligence,26*, pages 1645–1649, 2004.

[4] CMU. CMU face data set. *(http://kdd.ics.uci.edu/databases/faces/faces.data.html)*.

[5] Asari Gottumukkal and V.K. An improved face recognition technique based on modular PCA approach. *Pattern recognition Lett*, 25:429–436, 2004.

[6] Alejandro F.Frangi Jian Yang, David Zhang and jing-yu Yang. Two-Dimensional PCA:A New Approach to Appearance-Based Face Representation and Recognition. *IEEE Transactions on pattern analysis and machine intelligence*, 26, 2004.

[7] I.T Jolliffe. Principal component analysis. *Springer-Verelag,Newyork*, 2002.

[8] J.Y.Yang J.Yang. From image vector to matrix:a strightforward image projection technique-IMPCA vs.PCA. *Pattern Recognition*, 35, 2002.

[9] S.Chen K.Tan. Adaptively weighted sub-pattern PCA for face recognition. *Neuro Computing*, 64:505–511, 2005.

[10] K.VijayKumar and A.Negi. SubXPCA and generalized feature partitioning approach to principal component analysis. *Pattern recognition*, 41:1398–1409, 2008.

[11] Atul Negi and Kadappagiri Vijaya Kumar. Novel approaches to principal component analysis of image data based on feature partitioning framework. *Pattern Recognition*, 2008.

[12] D.G.Stork R.O.Duda, P.E.Hart. Pattern classification, second ed. *Wiley,New York*, 2004.

[13] Y.Zhu S.Chen. Sub-pattern based principal component analysis. *Pattern recognition*, pages 1081–1083, 2004.

[14] Yale. Yale face data set. *(http://cvc.yale.edu/projects/yalefaces/yalefaces.html).*