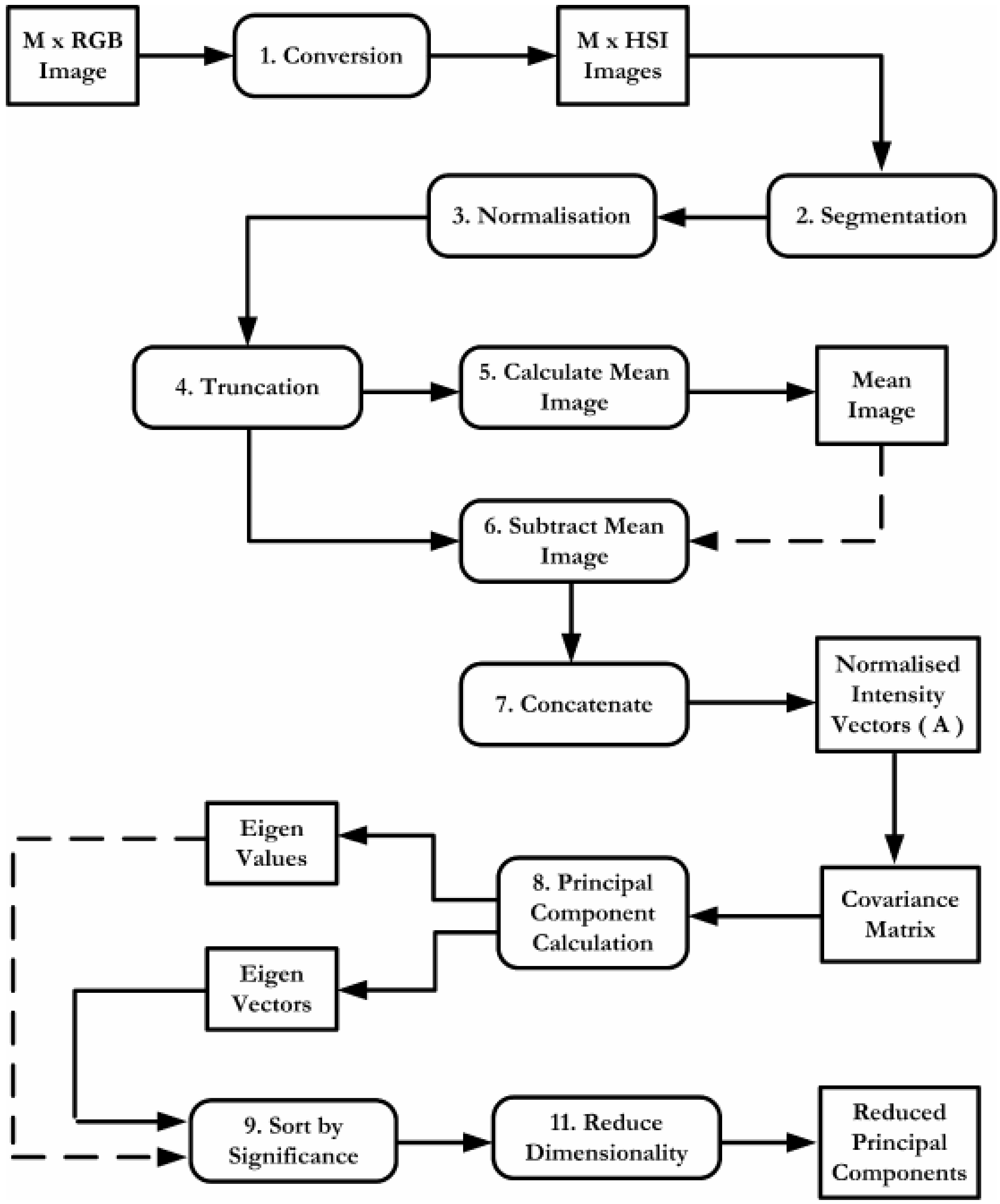
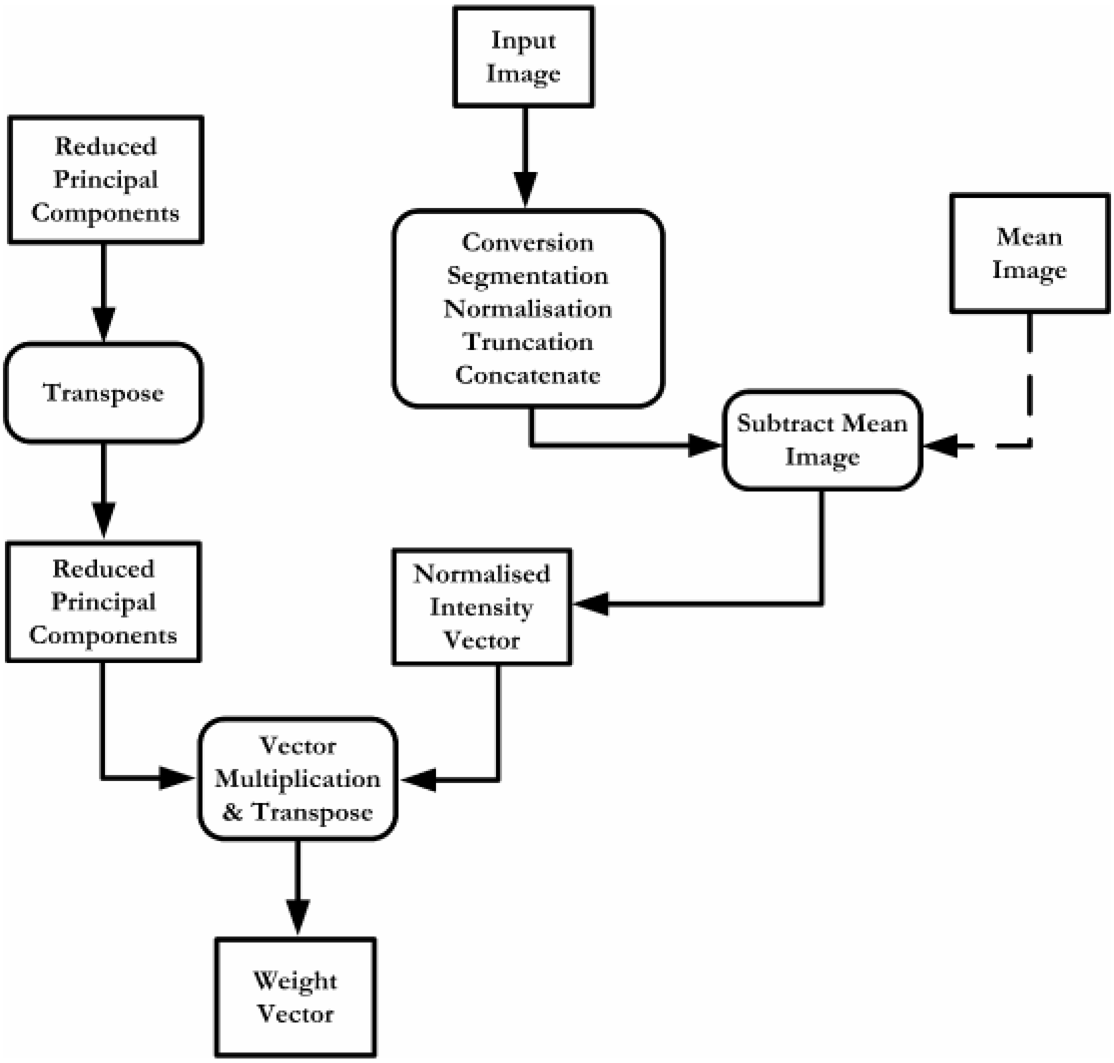
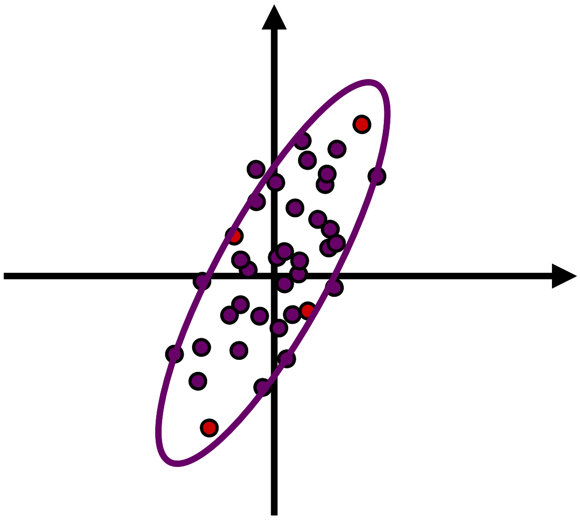
Proposed Framework for Training PCA on Hand Images



Proposed Framework for Classification of Hand Images with PCA



Example



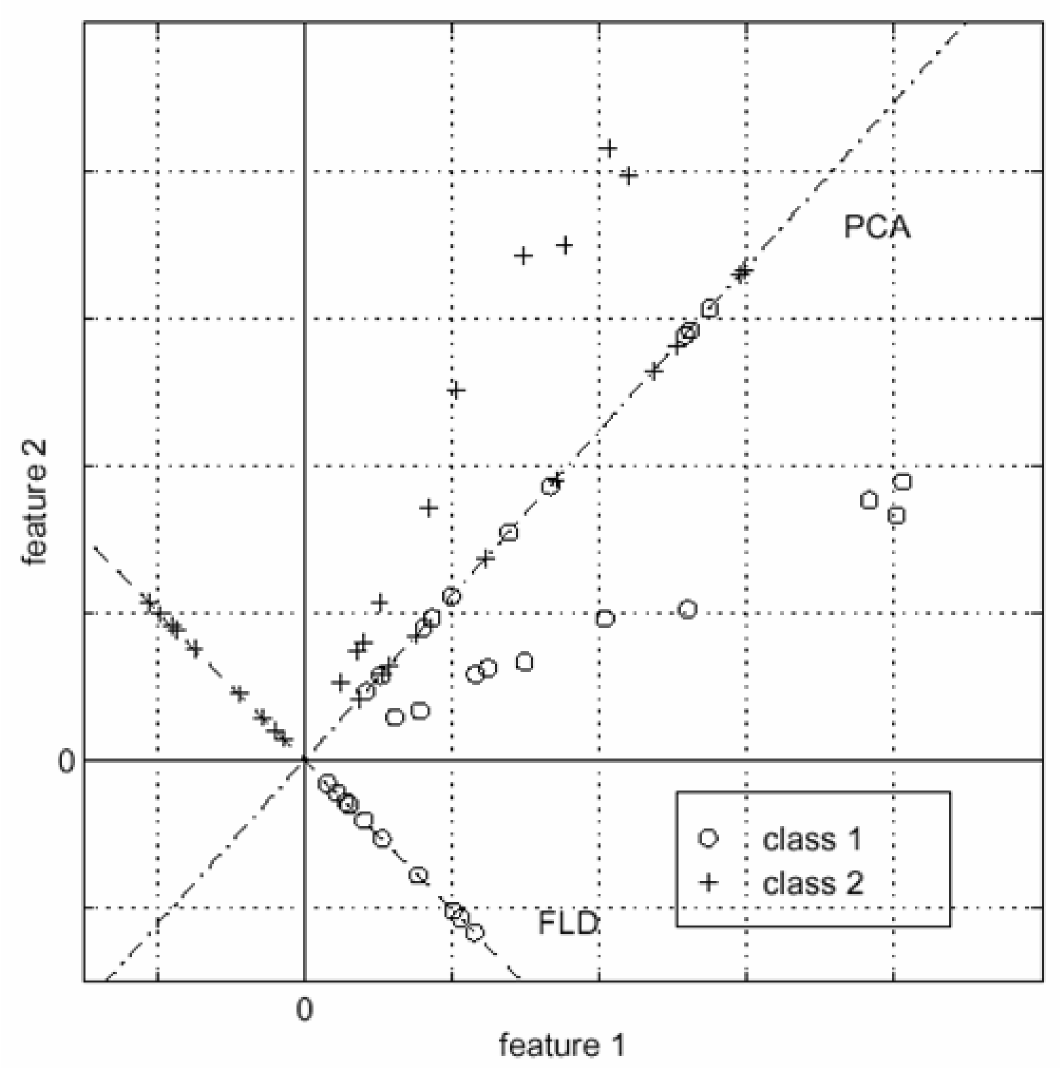
$$\mathbf{X} = \begin{bmatrix} 2 & -2 & 1 & -1 \\ 5 & -5 & -1 & 1 \end{bmatrix}$$

Eigenvalues-eigenvectors

$$eigenvalues = \begin{pmatrix} 15 & 0 \\ 0 & 0.5 \end{pmatrix} \qquad eigenvectors = \begin{pmatrix} .37 & .93 \\ -.93 & .37 \end{pmatrix}$$

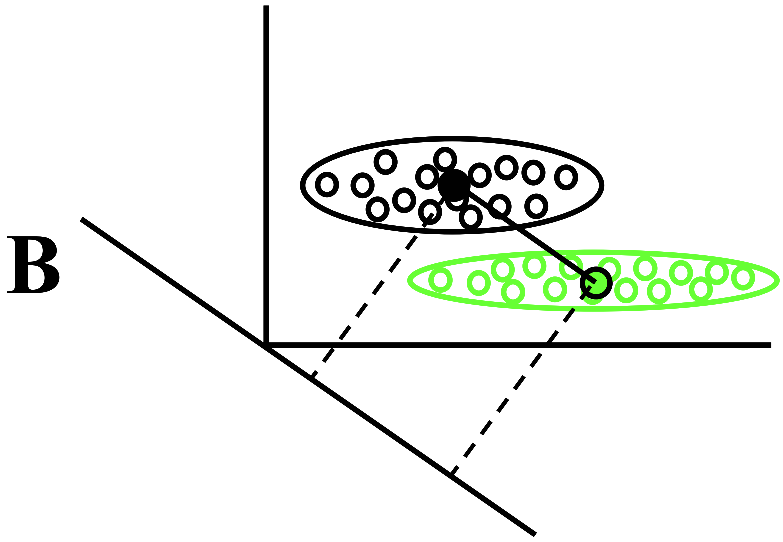
PCA does not always find the best combination to reveal the structure of data as it maximizes the overall variance and does not take into account the within class (nor between class) data distribution.

- No information about data structure, class labels, is used in PCA.



Belhumeur(97)

How to best separate the following data clusters ?



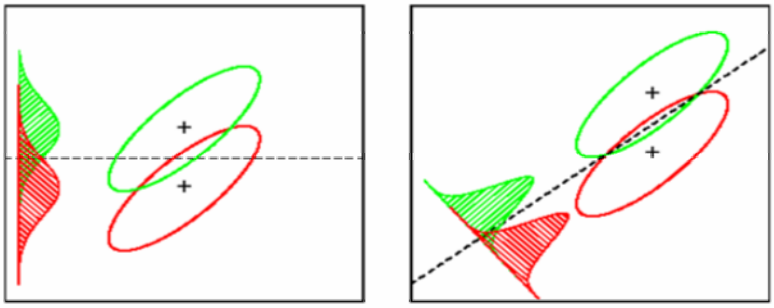
The maximum distance between the mean values of the cluster lies in B direction defines line passing through 0.

- The value of projection $Y=B^T \cdot X$
- If data clusters are small B is a good choice; find more projection directions in space orthogonal to direction B.

If we denote X_1 the set of point from class 1 and X_2 the set of points for class 2, the direction of B is:

$$B = \frac{\bar{X}_1 - \bar{X}_2}{\left\| \bar{X}_1 - \bar{X}_2 \right\|}$$

- This is not the optimal solution.
- Covariance of the two distributions should be taken into account:



Elements of statistical learning (Hasti, Tibshirani and Friedman 2001)

Goal

Find transformation $Y = W^T X$ that maximizes the distance between the projected mean values while keeping the within class variances low to separate the classes:

$$\left|\bar{Y}_1 - \bar{Y}_2\right| = \left|W^T\left(\bar{X}_1 - \bar{X}_2\right)\right|$$

- It can be written as:

$$\left|\bar{Y}_1 - \bar{Y}_2\right|^2 = \left(\bar{Y}_1 - \bar{Y}_2\right)\left(\bar{Y}_1 - \bar{Y}_2\right)^T$$

$$= \left(W^T\left(\bar{X}_1 - \bar{X}_2\right)\right)\left(W^T\left(\bar{X}_1 - \bar{X}_2\right)\right)^T$$

$$= W^T\left(\bar{X}_1 - \bar{X}_2\right)\left(\bar{X}_1 - \bar{X}_2\right)^T W$$

$$= W^T S_B W$$

$\boldsymbol{S_B}$ is the between-class scatter matrix

- The distance should be large relatively to the variance (scatter) to separate the classes

We would like to find the transform W that maximizes the determinant of the between classes scatter while minimizing the determinant of the within classes scatter.

- Move images of the same hand signs closer together, while moving images of difference hand signs further apart.

Reminder: PCA chooses projection to maximize determinant of total scatter:

$$W_{opt} = \arg\max_W \left|W^T S_T W\right|$$

Total scatter:

$$S_T = \sum_k^N (x_k - \mu)(x_k - \mu)^T$$

Here:

$$W_{opt} = \arg\max_{W} \frac{\left|W^T S_B W\right|}{\left|W^T S_W W\right|}$$

A vector maximizing the above ratio must satisfy the generalized eigenvalue problem:

$$S_B w_i = \lambda_i S_W w_i \quad i = 1,2,\ldots,m$$

If the matrix S_W is non-singular, it reduces to the eigenvalue problem. Find λ and w such as:

$$S_W^{-1} S_B w_i = \lambda_i w_i$$

For a set of *N* sample images:

$$\{X_1, X_2, \ldots, X_c\}$$

With each image belonging to one of the c classes of the dataset:

$$\{X_1, X_2, \ldots, X_c\}$$

We can define the between-class scatter matrix:

$$S_B = \sum_{i=1}^{c} P_i (\mu_i - \mu)(\mu_i - \mu)^T$$

and the within-class scatter matrix:

$$S_W = \sum_{i=1}^{c} P_i \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T$$

- Reminder: $S_T = S_B + S_w$

- "In the LD feature space, classification scheme such as the nearest neighbour rule works more effectively" (CUI 95)

- Better space in terms of classification of signs

The Dimension problem

The discriminant analysis procedure encounters problem as soon as the the within-class scatter matrix becomes degenerate.

- High dimension of the input image compared to the reduced number of training samples ($N < n$)
- Number of classes <= Number of training samples
- Dimension of S_W is n*n

Need first to reduce the space dimension so as to obtain a within-class scatter matrix S_W which is not singular.

- Use PCA approach first:
  - Retain the variance
  - Eliminate meaningless (close to zero) eigenvalues
  - Projection to a lower-dimensional space
- Use LDA afterwards

The problem is now:

Find linear transformation W mapping n-dimensional image space into m-dimensional feature space ($m<n$):

$$y_k = W^T x_k \quad k = 1,2,\ldots,N \quad y_k \in \Re^m$$

with:

$$W_{optim}^T = W_{lda}^T W_{pca}^T$$

$$W_{pca} = \arg\max_W \left| W^T S_T W \right|$$

$$W_{lda} = \arg\max_W \frac{\left| W^T W_{pca}^T S_B W_{pca} W \right|}{\left| W^T W_{pca}^T S_W W_{pca} W \right|}$$

Algorithm 22.6: Canonical variates identifies a collection of linear features that separating the classes as well as possible.

Assume that we have a set of data items of g different classes. There are n_k items in each class, and a data item from the k'th class is $\boldsymbol{x}_{k,i}$, for $i \in \{1, \ldots, n_k\}$. The j'th class has mean $\boldsymbol{\mu}_j$. We assume that there are p features (i.e. that the $\boldsymbol{x}_i$ are p-dimensional vectors).

Write $\overline{\boldsymbol{\mu}}$ for the mean of the class means, i.e.

$$\overline{\boldsymbol{\mu}} = \frac{1}{g}\sum_{j=1}^{g} \boldsymbol{\mu}_j$$

Write

$$\mathcal{B} = \frac{1}{g-1}\sum_{j=1}^{g}(\boldsymbol{\mu}_j - \overline{\boldsymbol{\mu}})(\boldsymbol{\mu}_j - \overline{\boldsymbol{\mu}})^T$$

Assume that each class has the same covariance Σ, which is either known or estimated as

$$\Sigma = \frac{1}{N-1}\sum_{c=1}^{g}\left\{\sum_{i=1}^{n_c}(\boldsymbol{x}_{c,i} - \boldsymbol{\mu}_c)(\boldsymbol{x}_{c,i} - \boldsymbol{\mu}_c)^T\right\}$$

The unit eigenvectors of $\Sigma^{-1}\mathcal{B}$ — which we write as $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_d$, where the order is given by the size of the eigenvalue and $\boldsymbol{v}_1$ has the largest eigenvalue — give a set of features with the following property:

- Projection onto the basis $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ gives the k-dimensional set of linear features that best separates the class means.

Bibliography

Peter N. Belhumeur, J. P. Hespanha, and David J. Kriegman. “*Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection*”. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 19, NO. 7, JULY 1997.

Y. Cui and J. Weng, ``Appearance-Based Hand Sign Recognition from Intensity Image Sequences," *Computer Vision and Image Understanding*, vol. 78, pp. 157-176, 2000

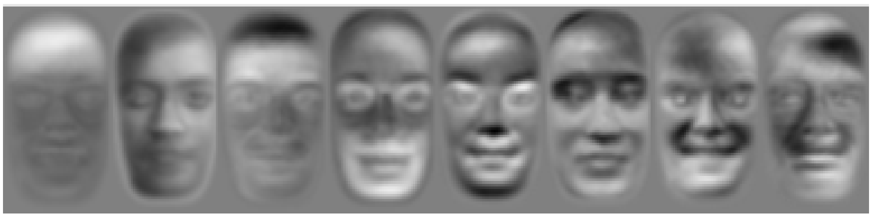
Daniel L. Swets, John (Juyang) Weng. “*Using Discriminant Eigenfeatures for Image Retrieval*”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 18, No 8 , pp. 831-836, 1996.

A. Martinez and A. Kak. “*PCA versus LDA*”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 23, No 2, pp. 228 -233, 2001.

A. Webb. *Statistical Pattern Recognition*, Arnold, 1999.

Eigenbase and Eigenfaces

(University of Delaware)



A few eigenfaces

"Illumination variations account for more in the dataset than variations due to changes in speaker faces"