Merging and Splitting Eigenspace Models

Tae-Kyun Kim
Senior Lecturer
https://labicvl.github.io/

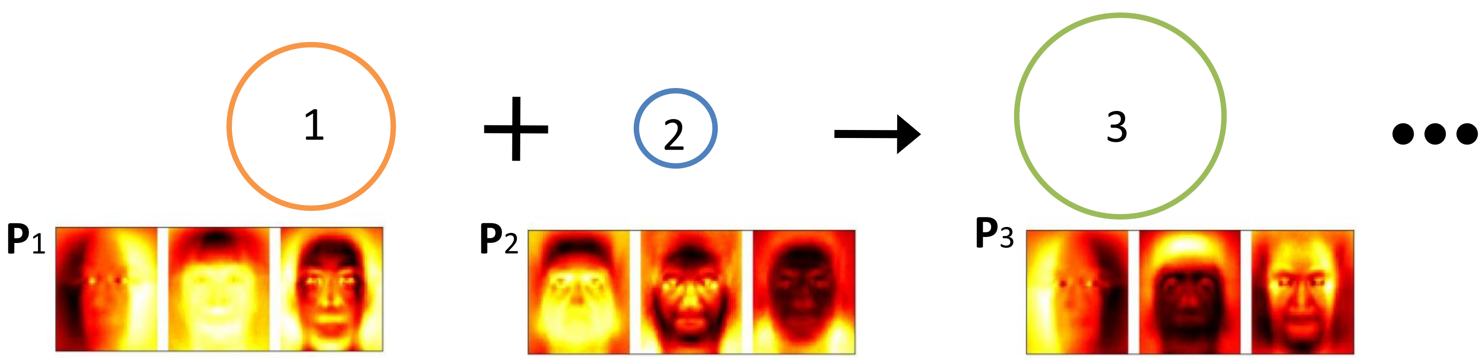
T-K. Kim, B. Stenger, J. Kittler and R. Cipolla, Incremental Linear Discriminant Analysis Using Sufficient Spanning Sets and Its Applications, International Journal of Computer Vision, 91(2):216-232, 2011.

Merging and splitting eigenspace models, P Hall, D Marshall, R Martin, IEEE Trans. on PAMI, 22 (9), 1042-1049, 2000.

Dynamically updating eigenspace

- Eigenspace models have a wide variety of applications, such as classification for recognition systems.
- In practice, we need to build the engenspace models for numerous images: those images may not be given all initially, but incrementally.

- Our goal is to dynamically update the eigenspace models, when new data entries are given or existing data points are removed.
- The mean also needs to be updated.



Incremental PCA

Batch vs Incremental

- In batch computation: all observations are used simultaneously to compute the eigenspace model.
- In incremental computation: an existing eigenspace model is updated using new observations.

Requirements: methods need to

- handle a change in the mean.
- add multiple new observations than exactly one observation at a time.

Pros and Cons

- Benefits: an incremental method
  - does not need all observations at once - thus, reducing storage requirements and making large problems feasible.
  - Even if all observations are available, is usually faster to compute a new eigenspace model by incrementally than by batch computation.
- Disadvantage: is their accuracy compared to batch methods. When only a few incremental updates are made, the inaccuracy is small.

Merging and splitting eigenspace models

– We learn a deterministic method that given two eigenspace models - each representing a set of *N*-dimensional observations - will:

1) Merge the models to yield a representation of the union of the sets,
2) Split one model from another to represent the difference between the sets.

Eigenspace models and notations

– For a set of *M* data vectors, $\mathbf{x} \in R^N$, the covariance matrix is

$$\mathbf{C} = 1/M \, \Sigma_{all\ \mathbf{x}} (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T$$

where $\boldsymbol{\mu}$ is the data mean.

– PCA decomposes the covariance matrix s.t.

$$\mathbf{C} \simeq \mathbf{P} \boldsymbol{\Lambda} \mathbf{P}^T$$

where **P**, **Λ** are the matrices containing the first *d* eigenvectors and eigenvalues.

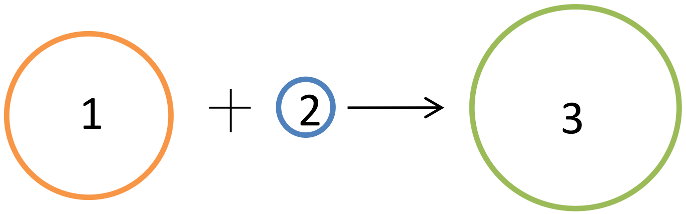
Incremental PCA

– Problem setting:
Input : given two sets of data represented by eigenspace models

$$\{\boldsymbol{\mu}_i, M_i, \mathbf{P}_i, \boldsymbol{\Lambda}_i\}_{i=1,2}$$

Output : compute the eigenspace model of the combined data

$$\{\boldsymbol{\mu}_3, M_3, \mathbf{P}_3, \boldsymbol{\Lambda}_3\}$$



– The combined mean is obtained as

$$\boldsymbol{\mu}_3 = (M_1\boldsymbol{\mu}_1 + M_2\boldsymbol{\mu}_2)/M_3$$

– The combined covariance matrix is

$$\mathbf{C}_3 = \frac{M_1}{M_3}\mathbf{C}_1 + \frac{M_2}{M_3}\mathbf{C}_2 + \frac{M_1 M_2}{M_3^2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T$$

where $\{\mathbf{C}_i\}$, $i=1,2$ are the covariance matrices of the first two sets and $M_3 = M_1 + M_2$.

Incremental PCA

– The eigenvector matrix **P**3 can be represented as

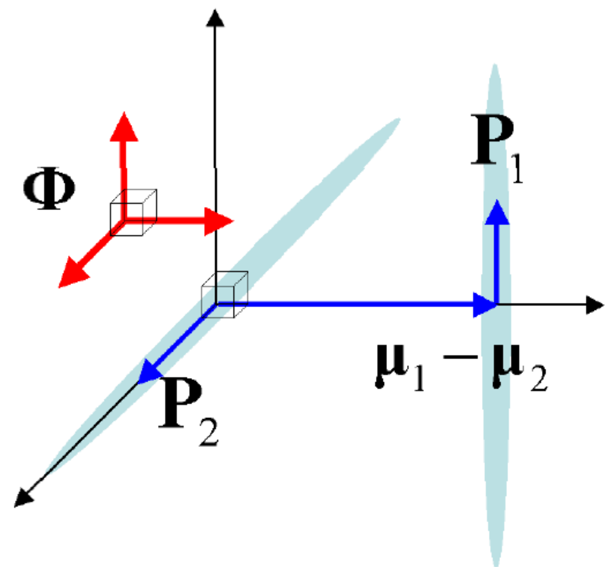
$$\mathbf{P}_3 = \mathbf{\Phi}\mathbf{R} = h([\mathbf{P}_1, \mathbf{P}_2, \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2])\mathbf{R}$$

where,

$\mathbf{\Phi}$ is the orthonormal matrix spanning the combined covariance matrix i.e. *the sufficient spanning set*,

R is a rotation matrix, and

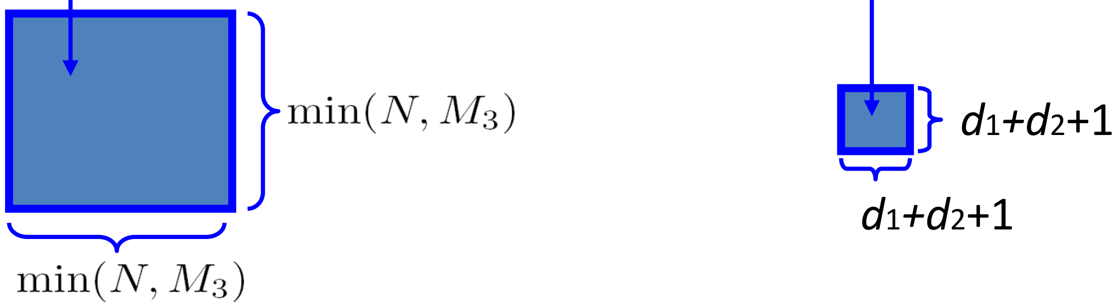
h is an orthonormalization function followed by removal of zero vectors.



Incremental PCA

– Using this representation, the eigenproblem is converted into a smaller eigenproblem as

$$\mathbf{C}_3 \simeq \mathbf{P}_3 \boldsymbol{\Lambda}_3 \mathbf{P}_3^T \quad \Rightarrow \quad \boldsymbol{\Phi}^T \mathbf{C}_3 \boldsymbol{\Phi} \simeq \mathbf{R} \boldsymbol{\Lambda}_3 \mathbf{R}^T$$



– By computing the eigendecomposition on the r.h.s. **Λ**3 and **R** are obtained as the respective eigenvalue and eigenvector matrices.

Incremental PCA

– The eigenvector matrix to seek is given as

$$\mathbf{P}_3 = \mathbf{\Phi R}$$

– Note the eigenanalysis on the r.h.s. only takes computations

$$O((d_1 + d_2 + 1)^3)$$

Where *d*1, *d*2 are the number of the eigenvectors stored in **P**1 and **P**2.

– The eigenanalysis in a batch mode on the l.h.s. requires

$$O(\min(N, M_3)^3)$$

Splitting eigenspace models

– Problem setting:

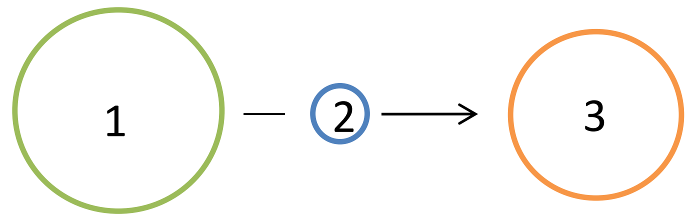
Input : given the first eigenspace model,

we remove the second from it,

$$\{\boldsymbol{\mu}_i, M_i, \mathbf{P}_i, \boldsymbol{\Lambda}_i\}_{i=1,2}$$

Output : to give the third model

$$\{\boldsymbol{\mu}_3, M_3, \mathbf{P}_3, \boldsymbol{\Lambda}_3\}$$



– Splitting means removing a subset of observations; the method is the inverse of merging in this sense.

– $M_3 = M_1 - M_2$.

– The new mean is: $\mu_3 = (M_1\mu_1 - M_2\mu_2)/M_3$

Splitting eigenspace models

– The new covariance matrix is

$$C_3 = M_1/M_3\, C_1 - M_2/M_3\, C_2 - M_2/M_1\, (\mu_2 - \mu_3)(\mu_2 - \mu_3)^T$$

– The eigenvector matrix $\mathbf{P}_3$ can be represented as $\mathbf{P}_3 = \mathbf{\Phi R} = \mathbf{P}_1\mathbf{R}$

where

$\mathbf{\Phi}$ is the orthonormal matrix spanning the new covariance matrix

i.e. *the sufficient spanning set*, and

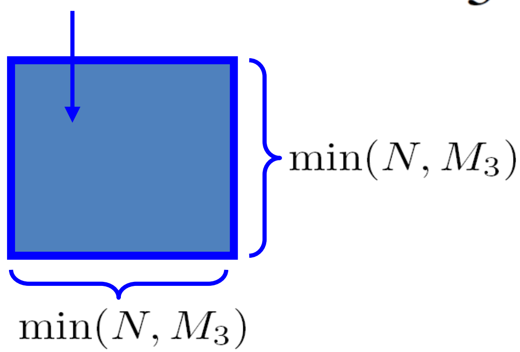
$\mathbf{R}$ is a rotation matrix.

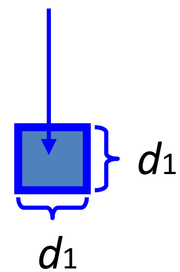
– It is impossible to regenerate information which was discarded when the overall model was created. Thus, if we split one eigenspace model from a larger one, the eigenvectors of the remnant must still form some subspace of the larger.

Splitting Eigenspace Models

– Using this representation, the eigenproblem is converted into a smaller eigenproblem as

$$\mathbf{C}_3 \simeq \mathbf{P}_3 \mathbf{\Lambda}_3 \mathbf{P}_3^T \quad \Longrightarrow \quad \mathbf{\Phi}^T \mathbf{C}_3 \mathbf{\Phi} \simeq \mathbf{R} \mathbf{\Lambda}_3 \mathbf{R}^T$$





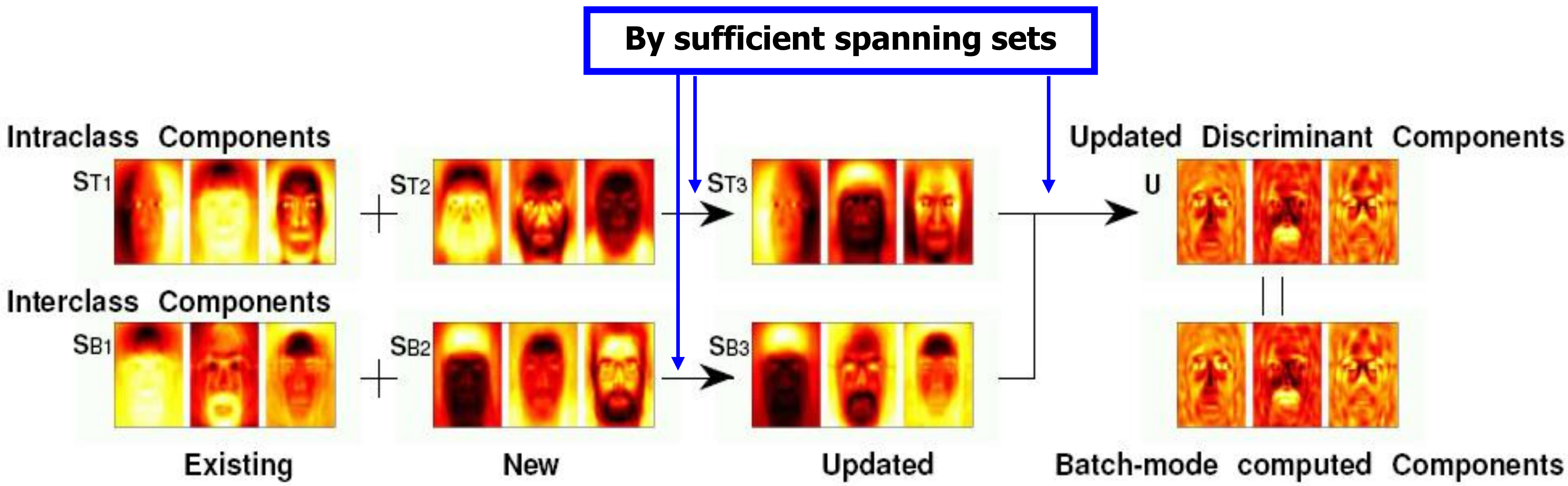
– By computing the eigendecomposition on the r.h.s. **Λ**3 and **R** are obtained as the respective eigenvalue and eigenvector matrices.

– The eigenvector matrix to seek is given as **P**3 = **ΦR = P**1**R**

Experiments

- Similarly, we can compute LDA (Linear Discriminant Anlysis) incrementally.
- We apply the *sufficient spanning set* approximation in each update step, i.e. for the between-class scatter matrix, the toal scatter matrix and the projected data matrix:

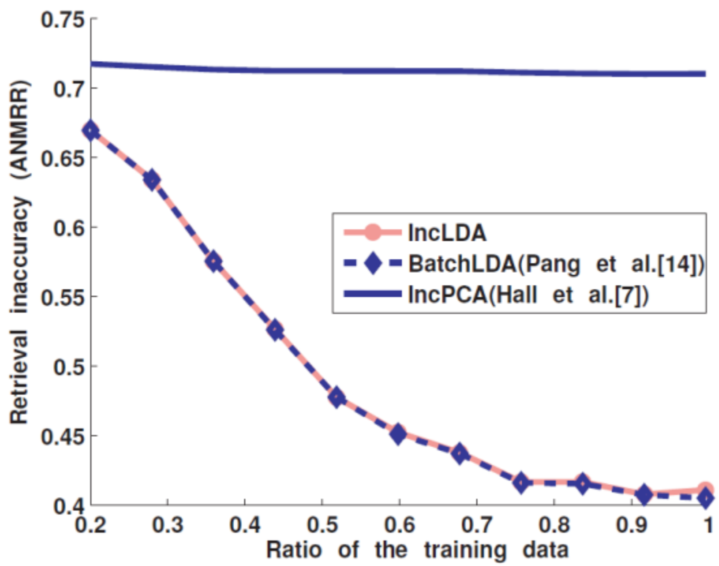
$$\max_{\arg \mathbf{U}} \frac{\mathbf{U}^T \mathbf{S}_B \mathbf{U}}{\mathbf{U}^T \mathbf{S}_W \mathbf{U}} = \boxed{\max_{\arg \mathbf{U}} \frac{\mathbf{U}^T \mathbf{S}_B \mathbf{U}}{\mathbf{U}^T \mathbf{S}_T \mathbf{U}}} \qquad \mathbf{S}_T = \mathbf{S}_B + \mathbf{S}_W$$

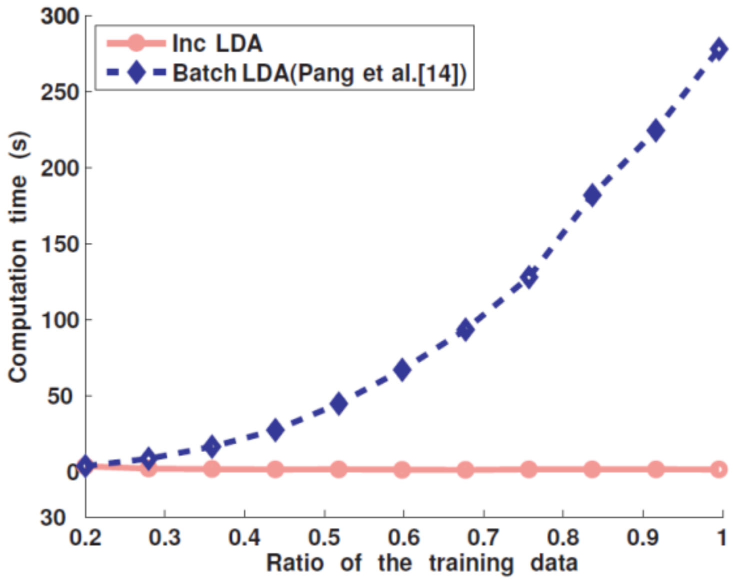


https://goo.gl/m6Mnov for questions

Experiments

MPEG7 face image datasets of 6370 images





https://goo.gl/m6Mnov for questions

Experiments

Caltech101 datasets (using BoW representations), up to 800 images per category