Face Recognition Using Laplacianfaces

He et al. (IEEE Trans PAMI, 2005)

presented by Hassan A. Kingravi

Overview

- Introduction
- Linear Methods for Dimensionality Reduction
- Nonlinear Methods and Manifold Learning
- LPP and its Connections to LDA and PCA
- Laplacianfaces for Face Analysis
- Experiments
- Conclusions

Introduction

- Face Recognition
  * Many methods developed
  * An example of the Appearance-Based Method, i.e. templates extracted from given information

- Current Approach (Laplacianfaces)
  * Builds on the dimensionality reduction approach
  * Takes into account the possibility of the distribution of the data on a non-linear subspace by preserving local structure (Locality Preserving Projections)
  * Generalizes to unseen points (problem in most nonlinear methods)

Dimensionality Reduction

- Images represented as vectors in extremely high-dimensional space

- Idea is that we can create decision boundaries between different classes of objects

- Dimensionality reduction is required because:
  a) Inherent structure of data may not be apparent in high dimensional space; vector can have irrelevant attributes that contain little useful information
  b) A projection makes the data more tractable to manage

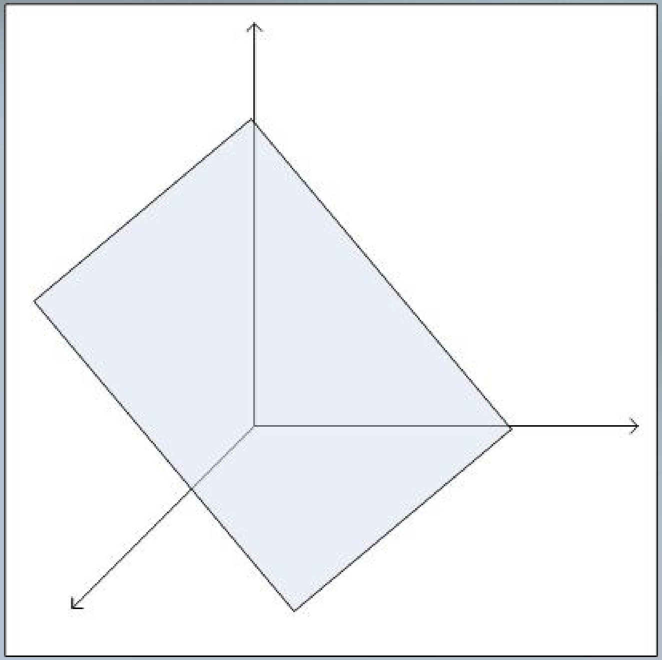
- Reduction can be linear (PCA, LDA) or nonlinear (ISOMAP)

DR Methods : PCA

- Projects high-dimensional data to low dimensional subspace
- Objective function maximizes variance globally:

$$\max_{w} \sum_{i=1}^{n} (y_i - \overline{y})^2$$

- Solution by finding the eigenvectors of the (reduced) covariance matrix $w_1, w_2, \ldots, w_k$.



DR Methods : PCA

- Since PCA maximizes global variance, it can be thought of re-representing as much of the original signal as possible.

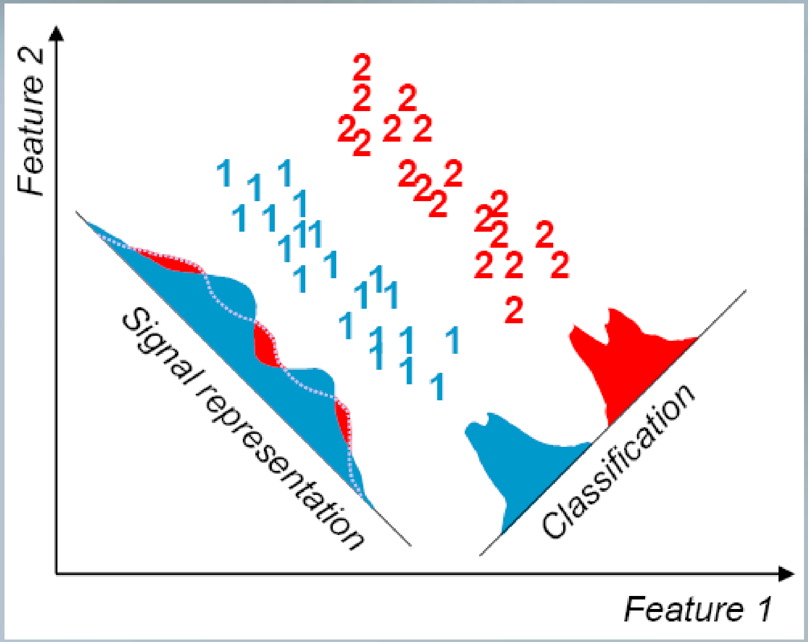
- This does NOT mean that the data is projected with an aim for classification; rather, it is compressed.

- Other methods exist which aim to project the vector based on multiclass discrimination.

DR Methods : LDA

- LDA seeks directions of projection that are efficient for discrimination.
- The objective function maximizes between-class scatter over within-class scatter:

$$\max_{w} \frac{w^T S_B w}{w^T S_W w}$$



DR Methods : LDA

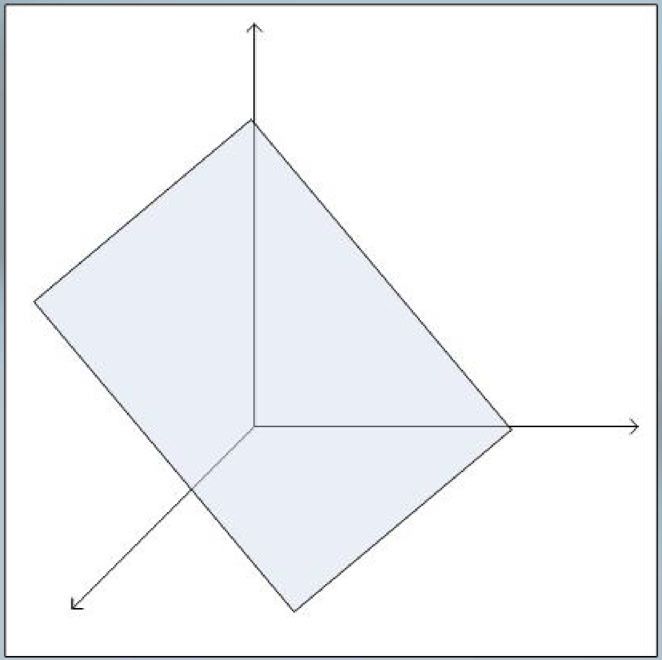
- LDA seeks directions of projection that are efficient for discrimination.

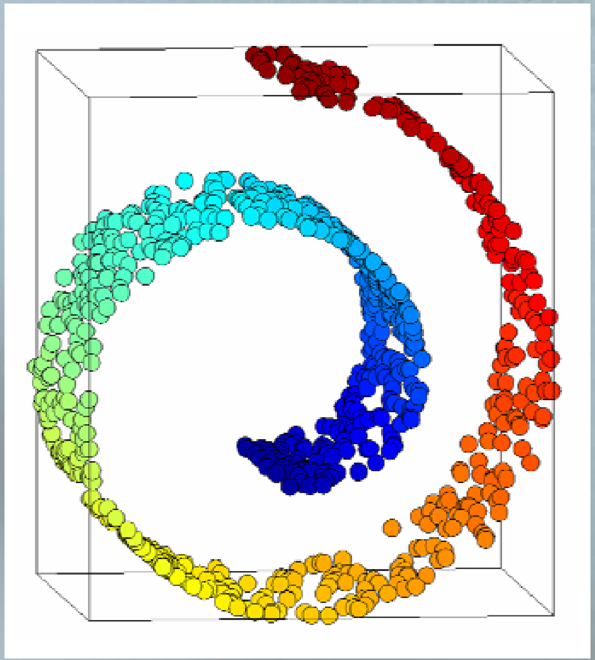
- LDA is defined in terms of the number of classes of projection; hence, it is a supervised learning algorithm (need to know class labels.) In the face recognition problem, one unique face represents a class.

- Good performance in general (better than PCA for faces), but still linear.

Nonlinear Methods

- When the linear subspace assumption is violated, for e.g. the figure on the right.





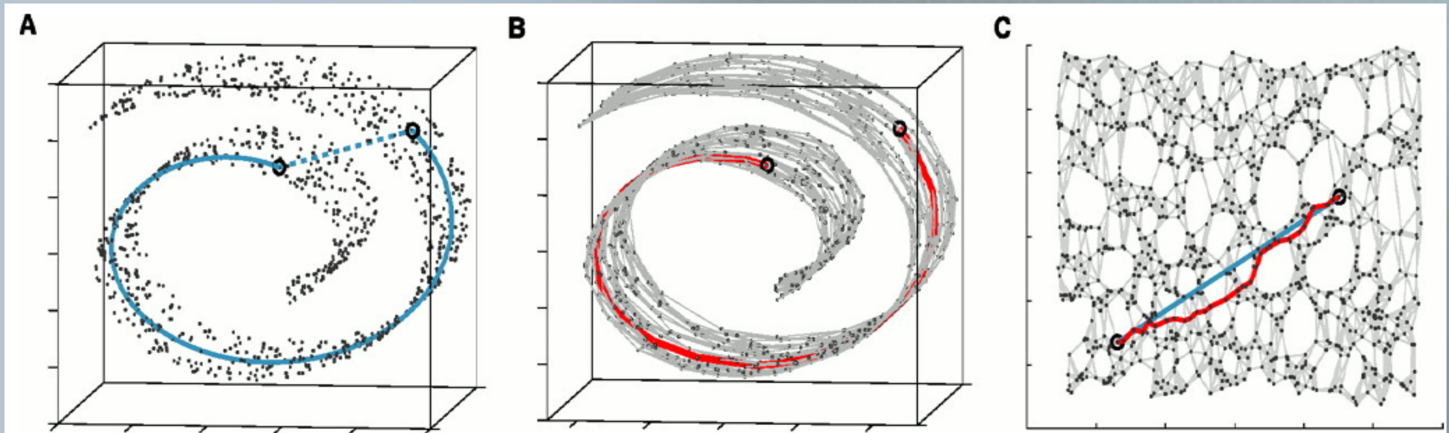
- The data lies on a nonlinear manifold, a mathematical space where the local area around a point may be Euclidean, but the overall structure may be more complicated

Nonlinear Methods

- Idea: manifolds arise naturally whenever there’s a smooth variation of parameters.
- Hypothesis; face recognition problems are non-linear in nature.
- Specifically, images that change smoothly over time (video).

Nonlinear Methods: ISOMAP

- An example of a nonlinear embedding method: ISOMAP.
- Need to see geodesic structure; solution: graph embeddings.
- Steps of ISOMAP:
  a) Find nearest neighbors to each sample (either use a k rule or a radius based on Euclidean distance) and construct a graph of the geodesic.
  b) Use Dijkstra's shortest path algorithm to find distances between all points.
  c) Apply multidimensional scaling.



[Tenenbaum et al., 1997]

Locality Preserving Projections

- Problems: ISOMAP is computationally intensive and the embedding's only defined on actual data points. Solution: Locality Preserving Projections, the method of Laplacianfaces.

- LPP is a linear method that approximates nonlinear methods (specifically, the Laplacian Eigenmap.)

- LPP minimizes the following objective function:

$$\min \sum_{ij} (y_i - y_j)^2 S_{ij}$$

where $S_{ij} = \exp(-\| x_i - x_j \|^2 / t), \ \| x_i - x_j \|^2 < \varepsilon$

Locality Preserving Projections

- The resulting mapping amounts to the following eigenvalue problem: $XLX^T w = \lambda XDX^T w$

- L is the Laplacian matrix, i.e. D – S, where S corresponds to the similarity values defined, and D is a column matrix which reflects how important a certain projection is. The more data points that surround a given point, the more "important" it is. Thus, the mapping preserves locality.

- The given equation corresponds to the Laplace-Beltrami operator on differential manifolds.

LPP and other methods

- LPP is a linear approximation to nonlinear methods, which takes locality into account.
- If one aims to preserve global structure only, let the neighborhood grow to infinity; data points are projected in directions of maximal variance, i.e. LPP becomes similar to PCA.
- LDA preserves discriminating information and global geometric structure; through manipulation, LPP can induce LDA. LDA is supervised, while LPP is unsupervised.

Laplacianfaces

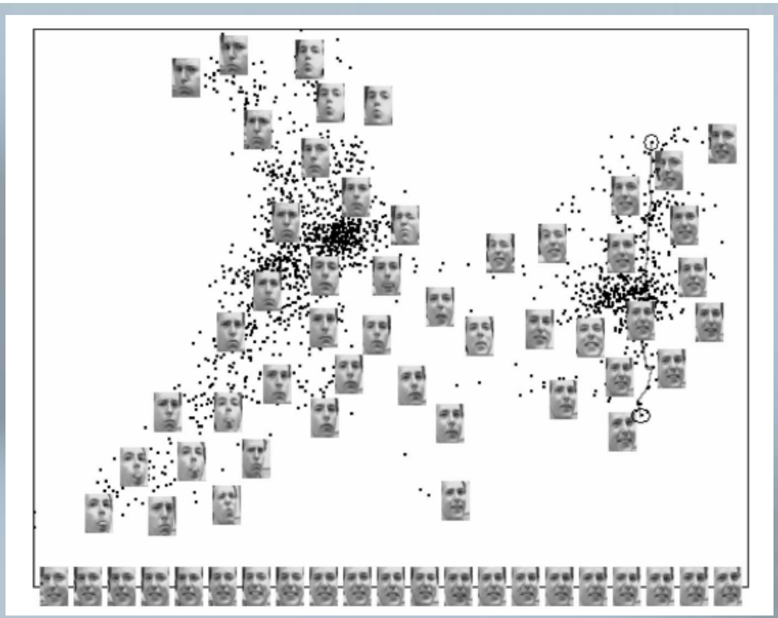
- Method:
  a) Because XDX^T can be sparse, the image is first projected onto a PCA subspace.

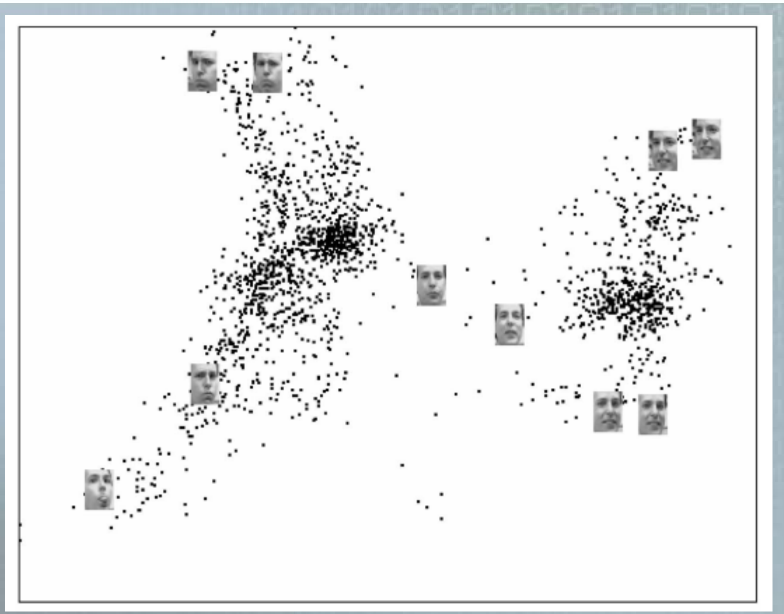
  b) The nearest neighbor graph is constructed, like ISOMAP. Laplacianfaces use the knn rule.

  c) Weights are chosen by the S_{ij} equation.

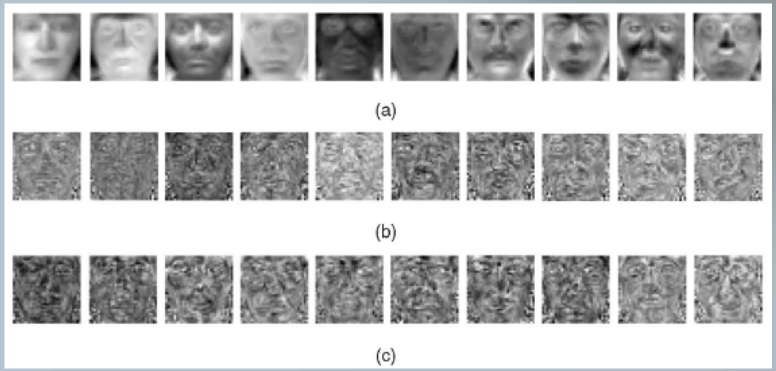
  d) The eigenmap $XLX^T w = \lambda XDX^T w$ is calculated. We get a set of k vectors that represent the new subspace.

Laplacianfaces





- Results from a mapping to a 2-d space.



Experimental Results

TABLE 2
Performance Comparison on the PIE Database

Approach	Dims	Error Rate
Eigenfaces	150	20.6%
Fisherfaces	67	5.7%
Laplacianfaces	**110**	**4.6%**

TABLE 3
Performance Comparison on the MSRA Database

Approach	Dims	Error Rate
Eigenfaces	142	35.4%
Fisherfaces	11	26.5%
Laplacianfaces	**66**	**8.2%**

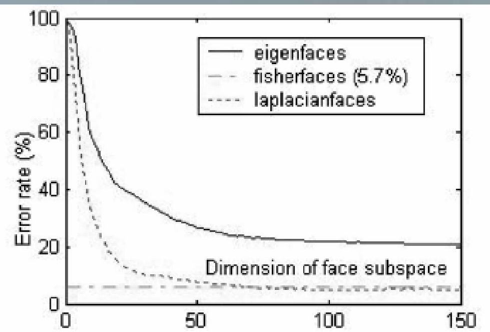


Fig. 8. Recognition accuracy versus dimensionality reduction on PIE database.

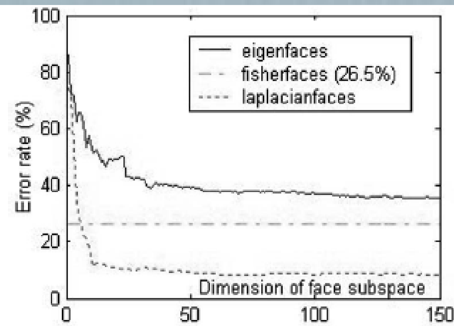


Fig. 10. Recognition accuracy versus dimensionality reduction on MSRA database.

Discussion & Conclusions

- Method has higher accuracy than both PCA and LDA approaches.
- LDA needs more than one sample per class to classify; LPP behaves like PCA.
- Method is faster than ISOMAP, and generalizes well (not specifically nonlinear.)
- LPP can also be applied to other machine learning issues.

References

- He et al. “Face Recognition Using Laplacianfaces.”
- Ricardo Gutierrez-Osuna, Pattern Recognition Slides, Fall 2006.