Visual Recognition: Faces

Tuesday, January 23

Why faces?

- Natural applications in human-computer interfaces (teleconferencing, assistive technology), organizing personal photos, surveillance,…
- Well-studied category, special structure
- We'll touch on a only a few general approaches

Faces

- Detection: given an image, where is the face?
- Recognition: whose face is it?

Ann

Image credit: H. Rowley

Challenges

- Face pose
- Occlusions
- Illumination
- Variable components (glasses, mustache, etc.)
- Differences in expression

Approaches

- Subspaces
  - e.g. Turk and Pentland, Belhumeur and Kreigman
- Shape and appearance models
  - e.g. Cootes and Taylor, Blanz and Vetter
- Boosting
  - e.g. Viola and Jones
- Neural networks
  - e.g. Rowley et al.
- SVMs
  - e.g. Heisele et al., Guo et al.
- HMMs
  - e.g. Nefian et al.

Eigenpictures/Eigenfaces

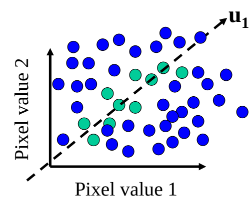
- Sirovitch and Kirby 1987: PCA to compress face images
- Turk and Pentland 1991: PCA + nearest neighbors to classify face images
- **Main idea**: face images are highly correlated; low-d subspace captures most appearance variation

Images as high-dimensional points

- Around d=80,000 pixels each
- To represent the space accurately, want num samples >> d
- But space of **face images** actually much smaller than space of all 80,000 dimensional images

PCA intuition

- Construct lower dimensional **linear** subspace that best explains variation of the training examples



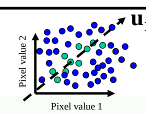
A face image
A (non-face) image

PCA

- N data points: $\mathbf{x}_1,\ldots,\mathbf{x}_N$ $\mathbf{x}_i$ in R^d
- Mean vector $\boldsymbol{\mu}$, covariance matrix Σ

What unit vector **u** in R^d captures the most possible variance of the data?

PCA



$$var(u) = \frac{1}{N}\sum_{i=1}^{N-1} \underbrace{\mathbf{u}^{\mathrm{T}}(\mathbf{x}_i - \mu)}_{\text{projection of data point}}(\mathbf{u}^{\mathrm{T}}(\mathbf{x}_i - \mu))^{\mathrm{T}}$$

$$= \mathbf{u}^{\mathrm{T}}\underbrace{\left[\sum_{i=1}^{N-1}(\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^{\mathrm{T}}\right]}_{\text{covariance of data points}}\mathbf{u}$$

$$= \mathbf{u}^{\mathrm{T}}\Sigma\mathbf{u}$$

Maximizing this is an eigenvalue problem→ use eigenvector(s) of Σ that correspond to the largest eigenvalue(s) as the new basis.

Eigenfaces

- Premise: Set of faces lie in a subspace of set of all images
- Use PCA to determine the *k* (*k*<*d*) vectors $\mathbf{u}_1,\ldots\mathbf{u}_k$ that span that subspace:

  $\mathbf{x} =\sim \boldsymbol{\mu} + w_1\mathbf{u}_1 + w_2\mathbf{u}_2 + \ldots + w_k\mathbf{u}_k$
- Then essentially use nearest neighbors in "face space" coordinates $(w_1,\ldots w_k)$ to do recognition

Eigenfaces

Training images:

$\mathbf{x}_1,\ldots,\mathbf{x}_N$

Eigenfaces

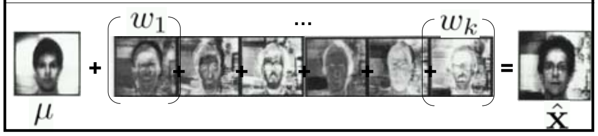
Top eigenvectors: $u_1, \ldots u_k$

Mean: μ

Eigenfaces

Face x in "face space" coordinates:

$$\mathbf{x} \rightarrow [\mathbf{u}_1^{\mathrm{T}}(\mathbf{x}-\mu), \ldots, \mathbf{u}_k^{\mathrm{T}}(\mathbf{x}-\mu)]$$
$$\rightarrow w_1, \ldots, w_k$$

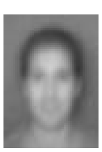


Eigenface recognition

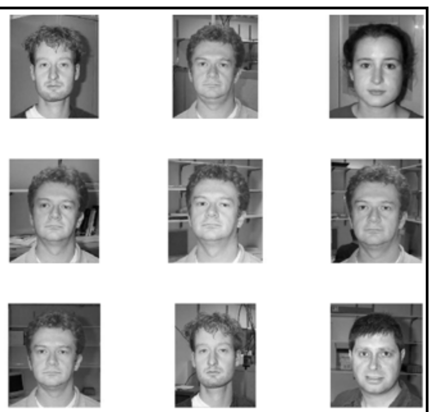
- Process labeled training images:
  - Run PCA
  - Project each training image onto subspace
- Given novel image:
  - Project onto subspace
  - If reconstruction error too large
    - Not a face
  - Else if too far from any training face
    - Unknown face
  - Else
    - Classify as closest training face in k-dimensional subspace

Small demo

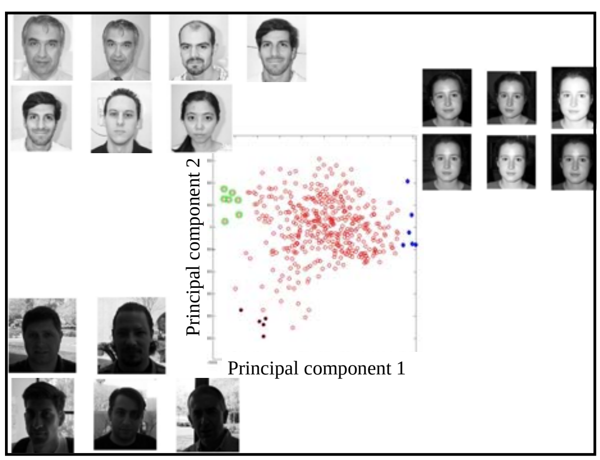
- Eigenfaces on the face images in the Caltech-4 database
- 435 images, same scale, aligned



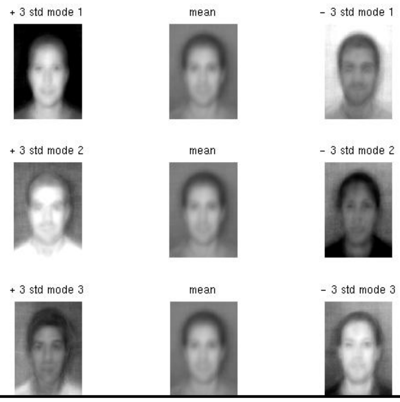
Mean



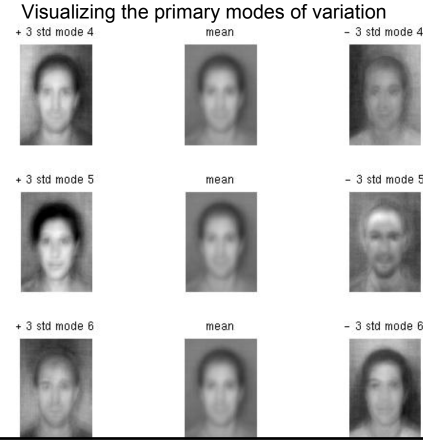
9 closest to mean



Visualizing the primary modes of variation



Visualizing the primary modes of variation



Clustering in the face subspace

Clustering in face subspace

Clustering in face subspace

Limitations

- PCA useful to *represent* data, but directions of most variance not necessarily useful for classification (see work by Belhumeur & Kreigman using LDA)
- Not appropriate for all data: PCA is fitting hyperplane to data / Gaussian where Σ is covariance matrix (see nonlinear techniques)
- In this application, assumptions about pre-processing applied to face images may be unrealistic
- Suited for what kinds of categories?

Fisherfaces

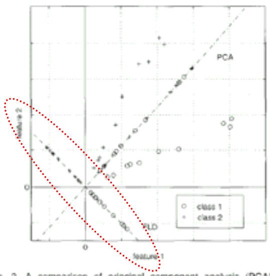


Fig. 2. A comparison of principal component analysis (PCA) and Fisher's linear discriminant (FLD) for a two class problem where data for each class lies near a linear subspace.

Belhumeur et al. PAMI 1997

Rather than maximize scatter of projected classes as in PCA, maximize ratio of between-class scatter to within-class scatter by using Fisher's Linear Discriminant

Non-linear dimensionality reduction

- Locally Linear Embedding (LLE), Roweis and Saul
- Isomap, Tenenbaum et al.
- Kernel PCA, Scholkopf et al.
- Laplacian Eigenmaps, Belkin and Niyogi

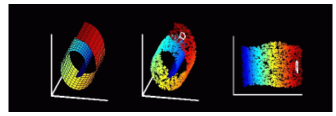


Image credit: Roweis and Saul

Active appearance models

- Eigenfaces model appearance only, and so cannot be robust to shape, pose and expression changes
- Active appearance models (Cootes and Taylor) model **shape** and **appearance**

Active appearance models

Factor out the faces' shape differences when comparing their texture / appearance

Coming up

- For Thursday: more on faces
  - Read Viola and Jones, and Sinha et al.
  - Review on Viola and Jones due
  - Zubair will present
- For Tuesday: part-based models
  - Read Felzenszwalb and Huttenlocher
  - Review due
  - Pushkala will present
  - Demo?