Image Analysis

PCA and Eigenfaces

Christophoros Nikou
cnikou@cs.uoi.gr

Images taken from:
- Computer Vision course by Svetlana Lazebnik, University of North Carolina at Chapel Hill.
- Computer Vision course by Michael Black, Brown University.
- Research page of Antonio Torralba, MIT.

Face detection and recognition

Detection  Recognition  “Sally”

Consumer application: iPhoto 2009

It can be trained to recognize pets!


http://www.apple.com/ilife/iphoto/

Consumer application: iPhoto 2009

iPhoto decides that this is a face
Outline

• Face recognition
  – Eigenfaces
• Face detection
  – The Viola and Jones algorithm.

The space of all face images

• When viewed as vectors of pixel values, face images are extremely high-dimensional
  – 100x100 image = 10,000 dimensions
• However, relatively few 10,000-dimensional vectors correspond to valid face images.
• Is there a compact and effective representation of the subspace of face images?

The space of all face images

• Insight into principal component analysis (PCA).
• We want to construct a low-dimensional linear subspace that best explains the variation in the set of face images.

Principal Component Analysis

• Given: $N$ data points $x_1, \ldots, x_N$ in $\mathbb{R}^d$
• We want to find a new set of features that are linear combinations of the original ones:
  \[ w(x_i) = u^T(x_i - \mu) \]
  ($\mu$: mean of data points)
• What unit vector $u$ in $\mathbb{R}^d$ captures the most variance of the data?

Principal Component Analysis

• The variance of the projected data:
  \[ \text{var}(w(x_i)) = \frac{1}{N} \sum_{i=1}^{N} w(x_i)w^T(x_i) = \frac{1}{N} \sum_{i=1}^{N} u^T(x_i - \mu)(u^T(x_i - \mu))^T \]
• We now estimate vector $u$ maximizing the variance:
  \[ u^T \Sigma u \]
subject to: \[ u^T u = |u|^2 = 1 \]
because any multiple of $u$ maximizes the objective function.
• The Lagrangian is \[ J(u, \lambda) = u^T \Sigma u + \lambda(1 - u^T u) \]
leading to the solution: \[ \Sigma u = \lambda u \]
which is an eigenvector of $\Sigma$. The one maximizing $J$ corresponds to the largest eigenvalue of $\Sigma$. 

References:

Principal Component Analysis (cont.)

- The direction that captures the maximum covariance of the data is the eigenvector corresponding to the largest eigenvalue of the data covariance matrix.
- The top $k$ orthogonal directions that capture the most variance of the data are the $k$ eigenvectors corresponding to the $k$ largest eigenvalues.

Eigenfaces: Key idea

- Assume that most face images lie on a low-dimensional subspace determined by the first $k$ ($k<d$) directions of maximum variance.
- Use PCA to determine the vectors or “eigenfaces” $u_1, ..., u_k$ that span that subspace.
- Represent all face images in the dataset as linear combinations of eigenfaces.

Eigenfaces example

Training images

Eigenfaces example (cont.)

Top eigenvectors

Mean: $\mu$

Eigenfaces example (cont.)

- Face $x$ in “face space” coordinates:
  \[
  x \rightarrow [u_1^T(x - \mu), u_2^T(x - \mu), ..., u_k^T(x - \mu)]^T
  = [w_1, w_2, ..., w_k]^T
  \]
  - Reconstruction:
  \[
  \hat{x} = \mu + w_1u_1 + w_2u_2 + \ldots + w_ku_k
  \]
Eigenfaces example (cont.)

- Any face of the training set may be expressed as a linear combination of a number of eigenfaces.
- In matrix-vector form:

$$E = \mu + \sum_{k=1}^{K} \lambda_k^k w_k^k$$

Error:

$$E^2 = \sum_{k=1}^{K} \lambda_k^k$$

Intra-personal subspace (variations in expression)

Extra-personal subspace (variations between people)

Face Recognition with eigenfaces

Process the training images:
- Find mean $\mu$ and covariance matrix $\Sigma$.
- Find $k$ principal components (eigenvectors of $\Sigma$) $u_1, \ldots, u_k$.
- Project each training image $x_i$ onto subspace spanned by principal components:

$$(w_{i1}, \ldots, w_{ik}) = (u_1^T(x_i - \mu), \ldots, u_k^T(x_i - \mu))$$

Given a new image $x$:
- Project onto subspace:

$$(w_{i1}, \ldots, w_{ik}) = (u_1^T(x - \mu), \ldots, u_k^T(x - \mu)).$$
- Optional: check the reconstruction error $x - \hat{x}$ to determine whether the new image is really a face.
- Classify as closest training face in $k$-dimensional subspace.


Computation of eigenvectors

- Given the $d$-dimensional vectorized images $x_i$ $(i=1, 2, \ldots, N)$, we form matrix $X=[x_1, x_2, \ldots, x_N]$ of size $d \gg N$.
- Covariance matrix:

$$\Sigma = \frac{1}{N} XX^T$$

- $\Sigma$ has dimensions $d \times d$ which, for example, for a 200x200 image is very large ($40000 \times 40000$).
- How do we compute its eigenvalues and eigenvectors?

Computation of eigenvectors (cont.)

- We create the $N \times N$ matrix:

$$T = \frac{1}{N} X^T X$$

- $T$ has $N$ eigenvalues ($\gamma_i$) and eigenvectors ($e_i$):

$$Te_i = \gamma_i e_i \iff \frac{1}{N} X^T X e_i = \gamma_i e_i \iff \frac{1}{N} \Sigma X e_i = \gamma_i X e_i$$

$$\iff \Sigma(X e_i) = \gamma_i(X e_i)$$

- $\gamma_i$ is an eigenvalue of $\Sigma$ and $X e_i$ is the corresponding eigenvector.

The importance of the mean image and the variations

Average of 100 of the images from the Caltech-101 dataset.

What is this?
The context plays an important role. Is it a car or a person? Both blobs correspond to the same shape after a 90 degrees rotation.

Reproduced from the web page of A. Torralba.

Average fashion model. Average of 60 aligned face images of fashion models.
Reproduced from the Perception Lab, University of St Andrews in Fife, UK.
http://perception.st-and.ac.uk

Application: MRF features for texture representation
- Each pixel is described by its $7 \times 7$ neighborhood in every $Lab$ channel.
- This results in a $7 \times 7 \times 3 = 147$-dimensional vector per pixel.
- Applying PCA to these vectors leads to a compact 6- to 8-dimensional representation of natural images.
- It captures more than 90% of the variation of the image pixels neighborhoods.

Application: view-based modeling
- PCA on various views of a 3D object.

Application: view-based modeling (cont.)
- Subspace of the first three PC (varying view angle).

Application: view-based modeling (cont.)

- View angle recognition.


Application: view-based modeling (cont.)

- Extension to more 3D objects with varying pose and illumination.


Application: Eigenheads

Principal components 1 and 2.

Principal components 3 and 4.

Application: Eigenheads (cont.)

Limitations

- Global appearance method
  - not robust to misalignment, background variation.

Limitations (cont.)

- Projection may suppress important detail
  - The smallest variance directions may not be unimportant. PCA assumes Gaussian data.

The shape of this dataset is not well described by its principal components.
**Limitations (cont.)**

- PCA does not take the discriminative task into account
  - Typically, we wish to compute features that allow good discrimination.
- Projection onto the major axis can not separate the green from the red data.
- The second principal component captures what is required for classification.

**Canonical Variates**

- Also called “Linear Discriminant Analysis”
- A labeled training set is necessary.
- We wish to choose linear functions of the features that allow good discrimination.
  - Assume class-conditional covariances are the same.
  - We seek for linear feature maximizing the spread of class means for a fixed within-class variance.

**Canonical Variates (cont.)**

- We have a set of vectorized images $x_i$ ($i=1,2,...,N$).
- The images belong to $C$ categories each having a mean $\mu_j$, ($j=1,2,...,C$).
- The mean of the class means:
  $$\mu = \frac{1}{C} \sum_{j=1}^{C} \mu_j$$

**Canonical Variates (cont.)**

- Between class covariance:
  $$S_B = \sum_{i=1}^{C} N_i (\mu_i - \mu)(\mu_i - \mu)^T$$
  with $N_i$ being the number of images in class $i$.
- Within class covariance:
  $$S_W = \sum_{i} \sum_{j} (x_{ij} - \mu_i)(x_{ij} - \mu_i)^T$$

Using the same argument as in PCA (and graph-based segmentation) we seek vector $u$ maximizing

$$\max_{u} \left\{ \frac{u^T S_B u}{u^T S_W u} \right\}$$

which is equivalent to

$$\max_{u} [u^T S_B u] \text{ s.t. } u^T S_W u = 1$$

The solution is the top eigenvector of the generalized eigenvalue problem:

$$S_B u = \lambda S_W u$$

---

- 71 views of 10 objects at a variety of poses at a black background.
- 60 images were used to determine a set of canonical variates.
- 11 images were used for testing.
• The first two canonical variates. The clusters are tight and well separated (a different symbol per object is used).
• We could probably get a quite good classification.