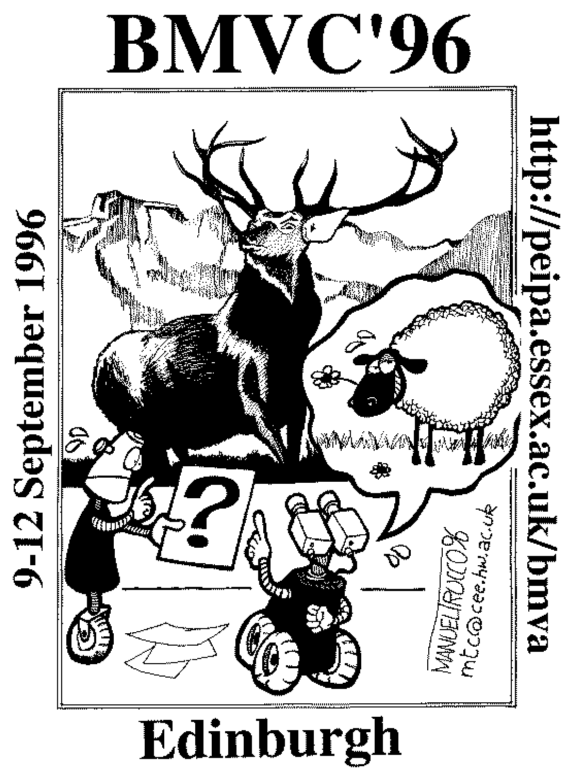
Lecture 30: Object Recognition II: Eigenspace Methods

Contents

© 2001–2008 Joachim Weickert

Introduction (1)



Poster of the 1996 British Machine Vision Conference illustrating the problem of object recognition. Author: E. Trucco (1996). Source: `http://www.ece.eps.hw.ac.uk/~mtc/tshirt.gif`.

Introduction

Goals

- For some given 2-D image, we want to check whether it represents a 3-D object in a database.
- The method should be general and also applicable to complicated geometric objects.
- Examples:
  – face recognition
  – recognition of different types of cars

Problems

- How can one represent a 3-D object by means of 2-D images?
- How can one find the most similar object?

Basic Idea

- Use 2-D images of the 3-D object that have been photographed under different directions and different illumination situations.
- Problem: requires a lot of disk space per object.
- Example: A 3-D object that is represented by bytewise coded 256×256 images which have been taken from 100 directions with 10 illumination variants requires 64 MByte.
- Can one reduce these huge memory requirements?

To this end we consider so-called *eigenspace methods (Eigenraumverfahren)* that are based on a principal component analysis (PCA, Hauptachsentransformation).

12 images of a 3-D object being viewed from different directions. Author: S. Kiefer (2006).

MIA 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

Eigenspace Representation

Assumptions

- Every image contains only a single object, and occlusions do not appear.
- The camera is sufficiently far away from the object such that we may approximate the nonlinear projective pinhole camera model by a linear affine one (cf. Lecture 26).
- All images are normalised in size, e.g. by ensuring in a face database that the image boundaries are given by the smallest rectangle that fully includes the image object.
- The image grey values are normalised such that they have zero mean, and their summed squared intensity is 1.
  Representing an image by a vector $\mathbf{f} = (f_1, ..., f_N)^\top$, this means that

$$\|\mathbf{f}\|^2 \ := \ \sum_{i=1}^{N} f_i^2 \ = \ 1.$$

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

A Simple Tool for Comparing Two Images

- Two normalised image vectors $\mathbf{f}$, $\mathbf{g}$ with zero mean are similar, if their inner product

$$\mathbf{f}^\top \mathbf{g} \;=\; \sum_{i=1}^{N} f_i\, g_i$$

  is close to 1.

- Sometimes this inner product is also called correlation (cf. Lecture 27).
  It ranges from -1 to 1.
  (Note that $\mathbf{f}$ and $\mathbf{g}$ have zero mean, and that $\|\mathbf{f}\|^2 = \|\mathbf{g}\|^2 = 1$.)

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

Principal Theorem of Eigenspace Representations

Consider m image vectors $\mathbf{f_1},...,\mathbf{f_m} \in \mathbb{R}^N$.
Usually one has less images than pixels, i.e. $m \ll N$.
Let $\bar{\mathbf{f}} := \frac{1}{m}\sum_{i=1}^{m} \mathbf{f_i}$ denote the average image.

Then the symmetric $N \times N$ *covariance matrix*

$$Q \;:=\; \frac{1}{m}\sum_{i=1}^{m} (\mathbf{f_i}-\bar{\mathbf{f}})\,(\mathbf{f_i}-\bar{\mathbf{f}})^\top$$

has at most m nonvanishing eigenvalues

$$\lambda_1 \geq ... \geq \lambda_m > 0$$

with corresponding orthonormal eigenvectors $\mathbf{v_1},...,\mathbf{v_m}$.
Every image $\mathbf{f_i}$ can be represented using these m eigenvectors of the covariance matrix:

$$\mathbf{f_i} \;=\; \bar{\mathbf{f}} + \sum_{j=1}^{m} a_{ij}\,\mathbf{v_j} \qquad (i = 1, ..., m)$$

with $a_{ij} := (\mathbf{f_i} - \bar{\mathbf{f}})^\top \mathbf{v_j}$.

Important Properties

Efficient Data Representation

- The average image $\bar{\mathbf{f}}$ has been subtracted, since one is only interested in deviations from the average shape.

- The covariance matrix Q resembles the structure tensor J from Lecture 19.
  Its eigenvectors point in the most characteristic directions of variation.
  This variation is measured by the corresponding eigenvalue.

- Usually, only k out of the m nonvanishing eigenvalues $\lambda_1,...,\lambda_m$ are *significantly* different from zero $(k \ll m)$.
  Thus we can approximate the m images $\mathbf{f_1},...,\mathbf{f_m}$ with only k eigenvalues of the covariance matrix:

$$\mathbf{f_i} \approx \bar{f} + \sum_{j=1}^{k} \alpha_{ij} \mathbf{v_j} \qquad (i = 1, ..., m).$$

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

- Since m represents the number of different directions and illumination variants, one has a much more compact representation:
  - Every convex combination of the inital images $\mathbf{f_1},...,\mathbf{f_m}$ can be approximated very well by using only $\mathbf{v_1},...,\mathbf{v_k}$.
  - Thus, the subspace spanned by $v_1,...,v_k$ is the space of learned 2-D views of the 3-D object.
  - Since the eigenvectors are orthogonal, this representation contains no redundancy.

Important Properties (3)

Efficient Distance Computation

- The Euclidian distance is closely connected to the correlation:
  For two normalised images $\mathbf{f_1}$, $\mathbf{f_2}$ with zero mean we have

$$
\begin{aligned}
\|\mathbf{f_1} - \mathbf{f_2}\|^2 &= \sum_{j=1}^{m} (f_{1j} - f_{2j})^2 \\
&= \sum_{j=1}^{m} \left(f_{1j}^2 - 2 f_{1j} f_{2j} + f_{2j}^2\right) \\
&= 1 - 2\,\mathbf{f_1}^\top \mathbf{f_2} + 1 \\
&= 2 - 2\,\mathbf{f_1}^\top \mathbf{f_2}.
\end{aligned}
$$

  Thus, minimising the distance $\|\mathbf{f_1} - \mathbf{f_2}\|$ is equivalent to maximising the correlation $\mathbf{f_1}^\top \mathbf{f_2}$.

Important Properties (4)

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

- Both can be approximated efficiently in the subspace spanned by $\mathbf{v_1}$, ..., $\mathbf{v_k}$, since this requires only $\mathcal{O}(k)$ instead of $\mathcal{O}(N)$ operations:

$$
\begin{aligned}
\|\mathbf{f_1} - \mathbf{f_2}\|^2 &\approx \left\| \bar{\mathbf{f}} + \sum_{j=1}^{k} \alpha_{1j} \mathbf{v_j} - \bar{\mathbf{f}} - \sum_{j=1}^{k} \alpha_{2j} \mathbf{v_j} \right\|^2 \\
&= \left\| \sum_{j=1}^{k} (\alpha_{1j} - \alpha_{2j})\, \mathbf{v_j} \right\|^2 \\
&= \sum_{j=1}^{k} (\alpha_{1j} - \alpha_{2j})^2
\end{aligned}
$$

  since $\mathbf{v_1}$, ..., $\mathbf{v_k}$ are orthonormal.

Important Shortcoming

- The initial images must be normalised.
  Already small perturbations may have a large impact on the vector representation of the image.
  Important e.g. for face recognition.

Computational Complexity

How Expensive is the Method?

- Often the covariance matrix Q is very large:
  Example: 256×256 images give 65536×65536 matrices.

- A direct computation of the eigenvalues and the eigenvectors of Q would be too time consuming.

Trick for Reducing the Complexity if $m \ll N$

- Consider the $N \times m$ matrix $D = (\mathbf{f_1} - \bar{\mathbf{f}}, ..., \mathbf{f_m} - \bar{\mathbf{f}})$.

- Instead of using the large $N \times N$ covariance matrix

$$Q = \frac{1}{m} D D^\top$$

  one considers the smaller $m \times m$ matrix

$$T = \frac{1}{m} D^\top D.$$

- One can show some very useful properties:
  - The m eigenvalues of T are also eigenvalues of Q.
    The remaining eigenvalues of Q are zero.
  - If $\mathbf{w_i}$ is eigenvector of T, then $\mathbf{v_i} := D\mathbf{w_i}$ is eigenvector of Q.

- If only a few, large eigenvalues of T and their eigenvectors are needed, simple numerical algorithms can be used (e.g. power method, method by von Mises).

Application to Object Recognition

Training Phase

- Create m normalised image vectors $\mathbf{f_1}$,...,$\mathbf{f_m}$ of a 3-D object using different direction and illumination conditions.
- Compute the average image $\bar{\mathbf{f}}$ and the covariance matrix Q.
- Compute the eigenvalues and eigenvectors of Q (via T).
- Use only the eigenvectors that correspond to the k eigenvalues that differ significantly from $\mathbf{0}$.
  The corresponding coefficient vectors in this subspace describe the object sufficiently well.

M I A

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

Recognition Phase

- Given: some image vector $\mathbf{g}$, for which we want to find the corresponding object in the data base.
- Normalise the image $\mathbf{g}$.
- Compare $\mathbf{g}$ with every 3-D object in the database by computing its representation in the corresponding eigenspace, and searching for the subspace in the database that represents $\mathbf{g}$ best.
- The error by representing $\mathbf{g}$ in a subspace spanned by k orthonormal vectors $\mathbf{v_1}$, ..., $\mathbf{v_k}$ can be measured by

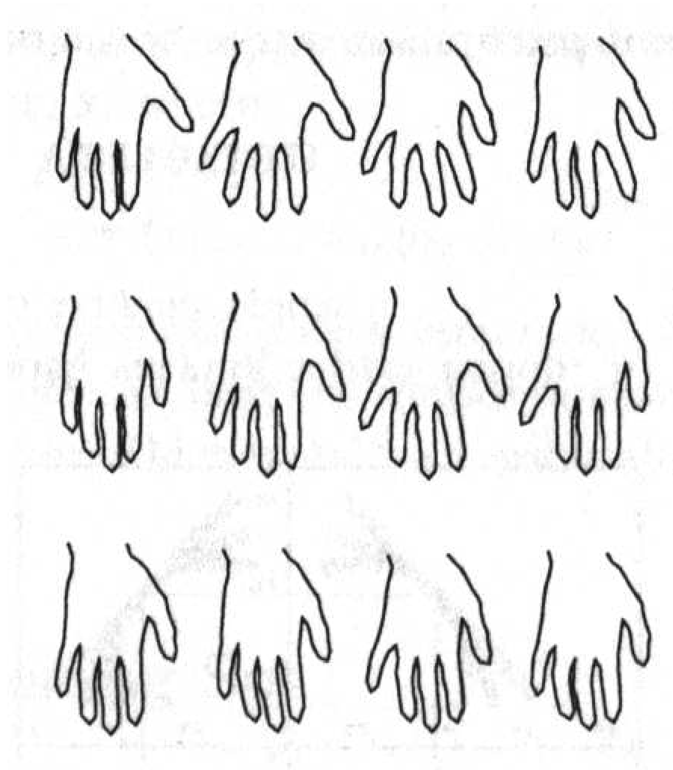
$$\left\| \mathbf{g} - \sum_{i=1}^{k} (\mathbf{g}^\top \mathbf{v_i}) \mathbf{v_i} \right\|^2$$

- The subspace that gives the smallest error represents the most similar object in the data base.

Application to Shape Variation

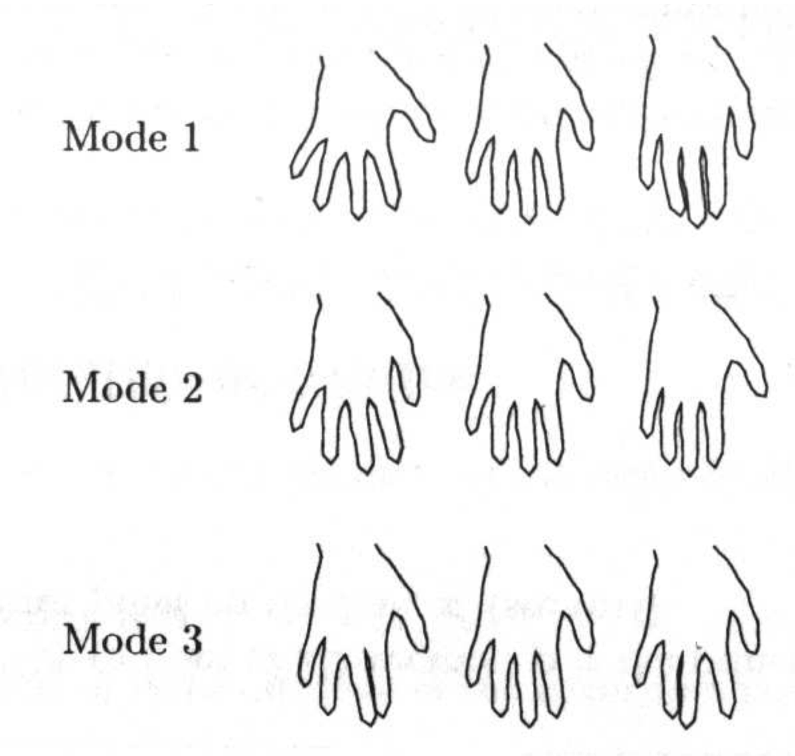
- Example: Organs in medical imaging may have very complicated shapes that one would like to describe with a few parameters only.
- Contours for instance can be described by Fourier coefficients (cf. Lecture 20) or spline coefficients.
  This vector of coefficients can also be used for an eigenspace representation.
- One learns the shape in eigenspace and considers only a small number of significant eigenvectors, e.g. $k = 4$.
- The corresponding eigenvalues λ_i measure the shape variation along the eigenvectors $\mathbf{v_i}$.
- If one considers $\bar{\mathbf{f}} + \alpha\,\mathbf{v_i}$ and varies the weight α in the inverval $[-2\lambda_i, 2\lambda_i]$, one gets a visual impression of the influence of the eigenvector $\mathbf{v_i}$.
- This shape variation along the eigenvectors gives the so-called *modes*.
- Such an *active shape model* can be used e.g. within flexible, knowledge based segmentation methods.

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24



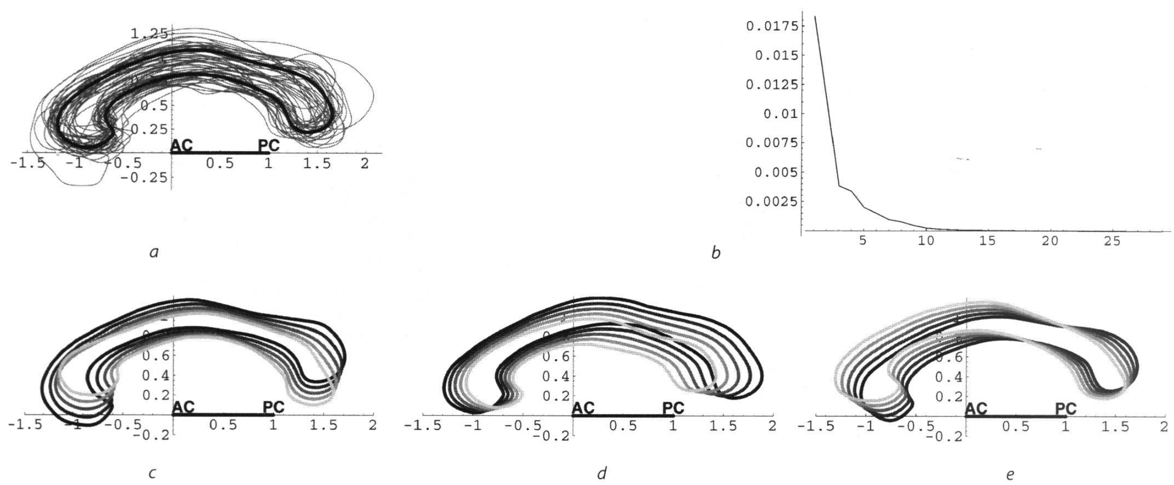
Training set for charactising the shape variation of hands. Authors: T. F. Cootes, C. J. Taylor (2001).

Application to Shape Variation (3)



Shape variation of hands. The first three modes are depicted, i.e. variations along the eigenvectors with the three largest eigenvalues are considered. Authors: T. F. Cootes, C. J. Taylor (2001).

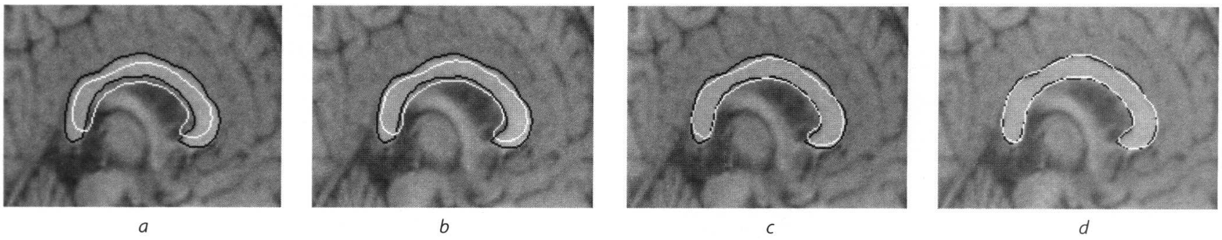
Application to Shape Variation (4)



Shape variation in a medical application. (a) Training set (corpus callosum). (b) Magnitude of the eigenvalues. (c),(d),(e) The first three modes. Authors: G. Székely, G. Gerig (2000).

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

Application to Shape Variation (5)



From left to right: Adaptation of the trained shape to the object by adding 1, 2, or 3 modes to the average shape. Knowing the average shape and the three most significant eigenvectors from a medical databas, only three parameters are required to describe a given shape almost perfectly. Authors: G. Székely, G. Gerig (2000).

Application to Shape Variation (6)

Shape variation in face recognition. The first two modes. Authors: T. F. Cootes, C. J. Taylor (2001).

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

Summary

- Image eigenspaces result from considering images as vectors and performing a principal component analysis of the covariance matrix.
- They are used in a training phase and a recognition phase for image data bases (e.g. for face recognition).
- very flexible tool that gives a compact representation, even for highly complicated shapes.
- requires normalisation of the input images

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

Literature

- E. Trucco, A. Verri: *Introductory Techniques for 3-D Computer Vision*. Prentice–Hall, Upper Saddle River, 1998.
  (This lecture is based on Section 10.4.)
- T. F. Cootes, C. J. Taylor:
  Statistical Models of Appearance for Computer Vision.
  Technical Report, Wolfson Image Analysis Unit, University of Manchester, England, 2001
  (`http://citeseer.ist.psu.edu/cootes04statistical.html`).
  (survey by two of the pioneers of that area)

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24