CS 534: Computer Vision
Appearance-based vision

Spring 2004
Ahmed Elgammal
Dept of Computer Science
Rutgers University

Outlines

- We will look into the major contributions in appearance-based vision
  - Appearance-based vision, problem definition and challenges
  - Subspace methods and PCA – review
  - Eigenfaces for face recognition
  - Parametric Appearance representations
  - Active shape and active appearance
  - Robust estimation and Eigen-tracking
  - Bilinear models and separation of style and content.

Appearance is important

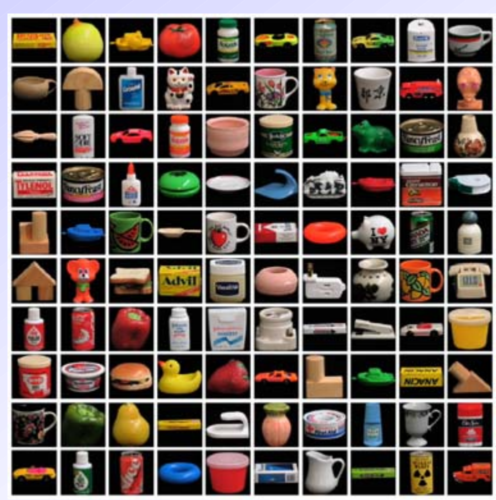
- 3D model-based recognition looks at object shape
- Is shape enough ?
- Vision deals with brightness images that are functions not only of shape but also intrinsic object and scene properties such as reflectance

- Representation of object appearance.

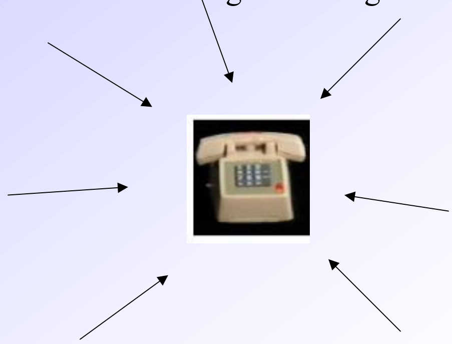
Appearance-based Recognition

How can we build representation of object appearance for recognition

Image from Shree K Nayar et al 1996 “RealTime 100 Object Recognition System”

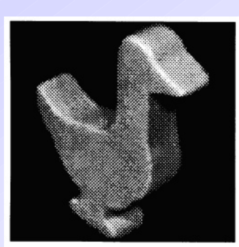


- Appearance is a function of the view point – object pose w.r.t. the camera
- We can collect many images of the object from many view points.
- How can we use these images to recognize the object



Why we need to do that:

- Shape is not enough. Object appearance is important in recognition
- Acquiring appearance models can be easier than acquiring 3D models



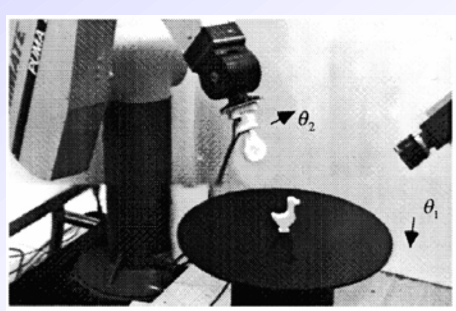
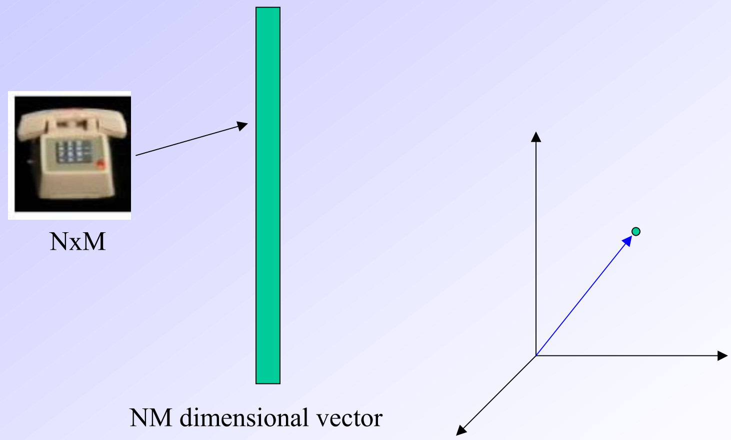


Figure from S. K. Nayar, et al, "Parametric Appearance Representation" 1996

- Different views is not the only possible variations
- Capture all possible variations
  - Object surface reflectance
  - Object pose
  - Illumination conditions
  - Sensor parameters
- Simply impractical
- For many applications the range of variations can be limited

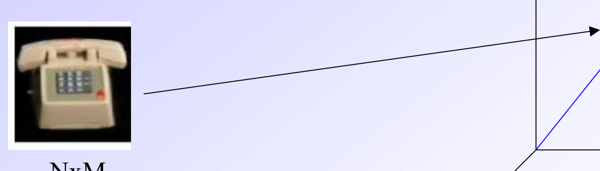
- The appearance of an object (rigid) is a combined effect of:
  - Its shape
  - Surface reflectance
  - Pose in the scene
  - Illumination condition

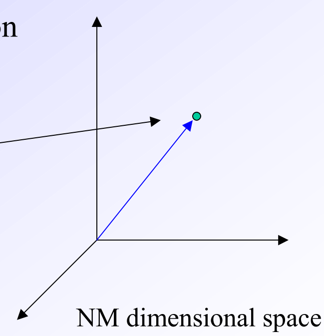
- Images are vectors in a high dimensional input space



Important Questions

- What is the relation between images of similar objects from the same view point.
- What is the relation between images of the same object from different view points ?
- … under different illumination
- They must be correlated





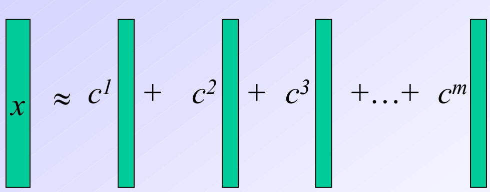
Subspace methods

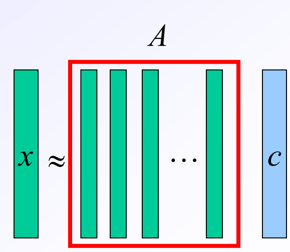
- Describe the images as linear combination of image basis
- Given a collection of points in a high dimensional space find a lower dimensional subspace to project these points into

Principle Component Analysis PCA

- Given a set of points $\{x_1, x_2, \cdots, x_N\}, x_i \in R^d$
- We are looking for a linear projection: a linear combination of orthogonal basis vectors

$$\underbrace{x}_{R^d} \approx A \cdot \underbrace{c}_{R^m, m << d}$$





What is the projection that minimizes the reconstruction error ? $E = \sum_i \| x_i - Ac_i \|$

Principle Component Analysis PCA

- Given a set of points

$$\{x_1, x_2, \cdots, x_N\}, x_i \in R^d$$

- Center the points: compute

$$\mu = \tfrac{1}{N}\sum_i x_i$$

$$P = [x_1 - \mu, x_2 - \mu, \cdots, x_N - \mu], x_i \in R^d$$

- Compute covariance matrix $Q = PP^T$
- Compute the eigen vectors for $Q \longrightarrow Qe_k = \lambda_k e_k$
- *Eigenvectors are the orthogonal basis we are looking for*

Singular Value Decomposition - Recall

- SVD: If **A** is a real *m* by *n* matrix then there exist orthogonal matrices

U ($m \times m$) and **V** ($n \times n$) such that

$\mathbf{U^t A V} = \Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_p)$ $\quad p = \min\{m,n\}$

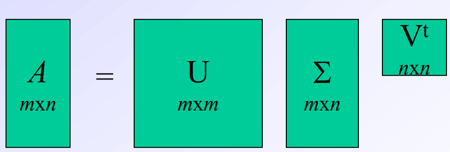
$\mathbf{U^t A V} = \Sigma \qquad \mathbf{A} = \mathbf{U}\Sigma\mathbf{V^t}$

- **Singular values**: Non negative square roots of the eigenvalues of $\mathbf{A^t A}$. Denoted σ_i, $i=1,\ldots,n$
- $\mathbf{A^t A}$ is symmetric $\Rightarrow$ eigenvalues and singular values are real.
- Singular values arranged in decreasing order.

$$A^t A = (U\sum V^t)^t (U\sum V^t) = V\sum^t U^t U \sum V^t = V\sum^t \sum V^t = V\sum^2 V^{-1}$$

$$(A^t A)V = V\sum^2$$

$$(A^t A)v = v\lambda$$



SVD for PCA

- SVD can be used to efficiently compute the image basis

$$PP^t = (U\Sigma V^t)(U\Sigma V^t)^t = U\Sigma V^t V \Sigma^t U^t = U\Sigma^t \Sigma U^t = U\Sigma^2 U^{-1}$$

$$(PP^t)U = U\Sigma^2$$

$$(PP^t)v = v\lambda$$

- U are the eigen vectors (image basis)

- Most important thing to notice: Distance in the eigen-space is an approximation to the correlation in the original space

$$\|x_i - x_j\| \approx \|c_i - c_j\|$$

PCA

$$R^d \longrightarrow \quad x \approx Ac \quad \longleftarrow R^m, m << d$$

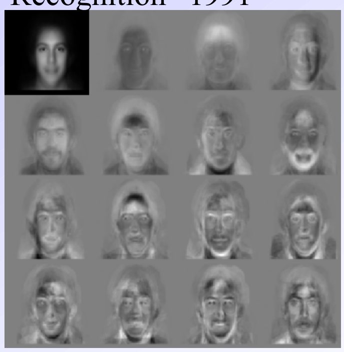
$$c \approx A^T x$$

- Most important thing to notice: Distance in the eigen-space is an approximation to the correlation in the original space

$$\|x_i - x_j\| \approx \|c_i - c_j\|$$

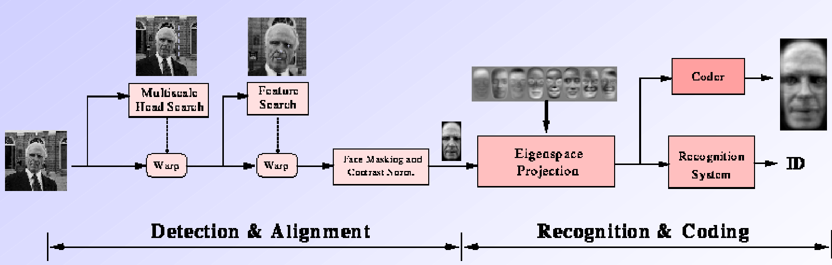
Eigenface

- Use PCA and subspace projection to perform face recognition
- How to describe a face as a linear combination of face basis
- Matthew Turk and Alex Pentland "Eigenfaces for Recognition" 1991



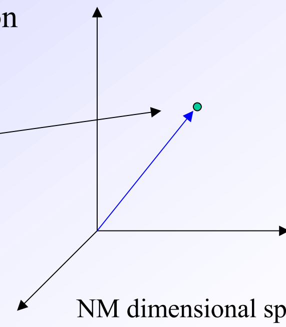
Face Recognition - Eigenface

- MIT Media Lab -Face Recognition demo page
  http://vismod.media.mit.edu/vismod/demos/facerec/



- What is the relation between images of similar objects from the same view point.
- What is the relation between images of the same object from different view points ?
- … under different illumination
- They must be correlated

NxM



NM dimensional space

Appearance Manifolds Learning

- Project all images to their eigen space
- Model each object view and illumination manifolds parametrically.

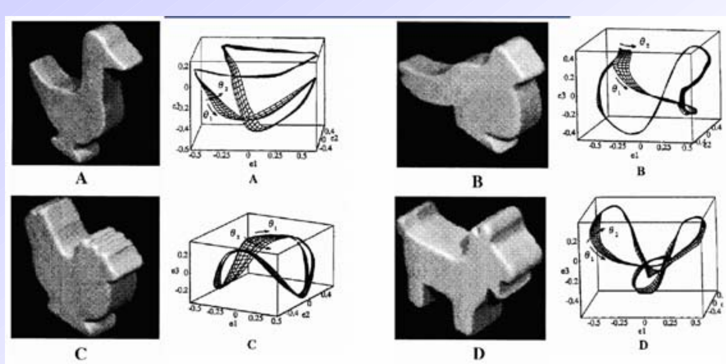


Figure from S. K. Nayar, et al, "Parametric Appearance Representation" 1996

Recognition

- Given a new image, segment and normalize
- Project into the eigen-space
- Find the closest manifold point
- Demo videos at:

http://www1.cs.columbia.edu/CAVE/research/publications/appearance_matching.html

Active shape – Active Appearance

- So far, our object are rigid
- Objective: model the shape/appearance of deformable objects
- Landmark-based approaches (e.g. Active shape/appearance models [Cootes et al 1995-])
- Deformation are modeled through linear models of certain landmarks through a correspondence frame.

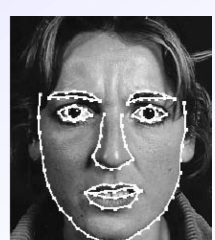
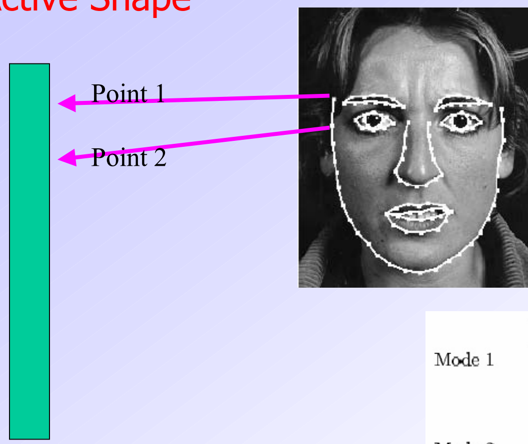
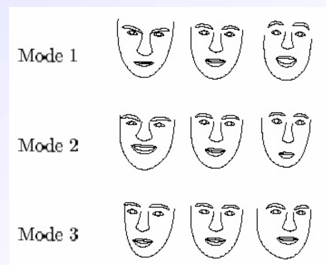



Figure from T. Cootes et al "Statistical models of appearance for Computer vision" 2000

Active Shape





x

$$x_i = A \cdot c_i$$

One vector for each image

Figure from T. Cootes et al "Statistical models of appearance for Computer vision" 2000

Active shape

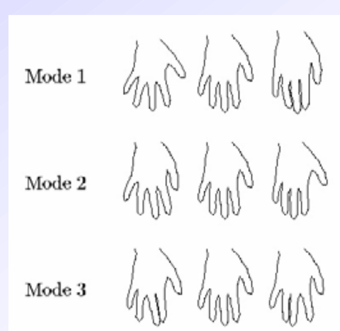


Figure from T. Cootes et al "Statistical models of appearance for Computer vision" 2000

Active Appearance

- Warp appearance (image batches) given a canonical shape to get rid of shape variations.

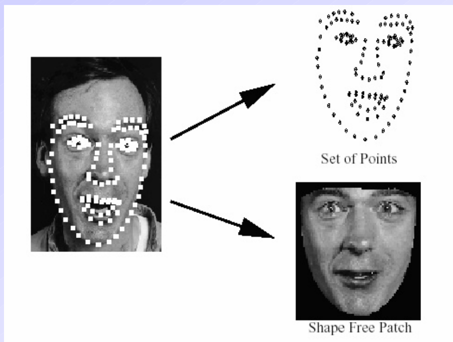
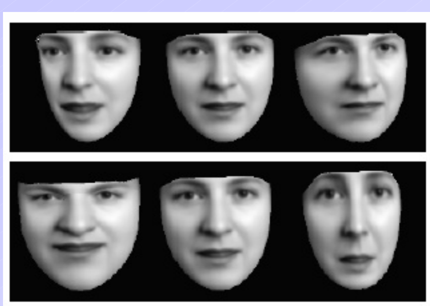
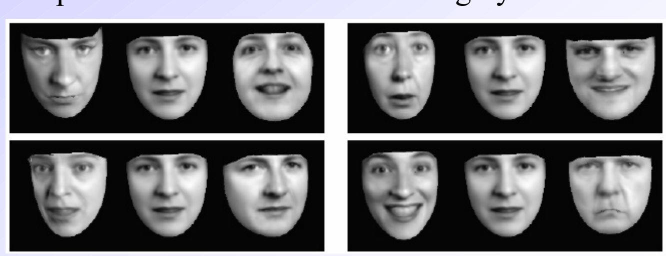


Figure from T. Cootes et al “Statistical models of appearance for Computer vision” 2000



2-shape modes

2-graylevel modes



4 – appearance modes (shape+graylevel)

Figure from T. Cootes et al “Statistical models of appearance for Computer vision” 2000

Robust Estimation and Eigen Reconstruction

- Michael J. Black and Allan D. Jepson " EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation"
- Use M-estimator for reconstruction

$$x \approx Ac$$

$$c \approx A^T x$$

$$E(c) = \sum_i \rho(x_i - Ac_i, \sigma)$$

Fig. 2. A simple example. *(a, b):* Training images. *(c, d):* Eigenspace basis images.

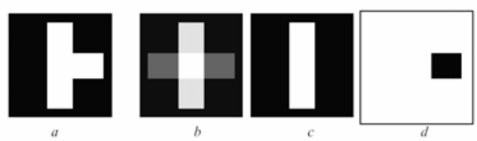
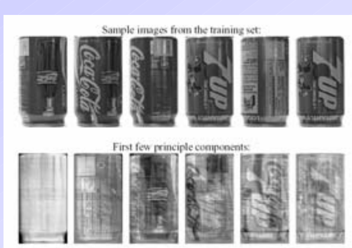


Fig. 3. Reconstruction. *(a):* New test image. *(b):* Least-squares reconstruction. *(c):* Robust reconstruction. *(d):* Outliers (shown as black pixels).

Figures from M. J. Black and A. D. Jepson " EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation" ECCV 1996

Eigen-tracking

- Michael J. Black and Allan D. Jepson " EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation"
- Formalize the tracking problem as a search for both eigenspace representation and image transformation



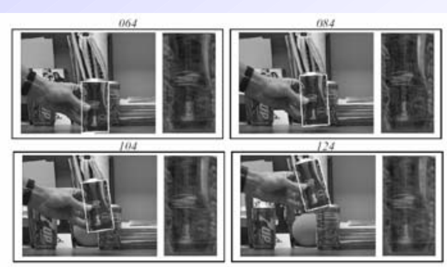
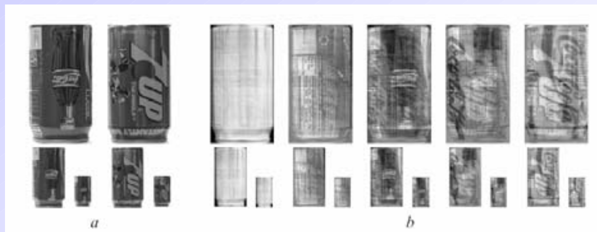


Fig. 7. Pickup Sequence. EigenTracking with translation and rotation in the image plane. Every 20 frames in the 75 frame sequence are shown.

Figures from M. J. Black and A. D. Jepson " EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation" ECCV 1996

Eigen tracking

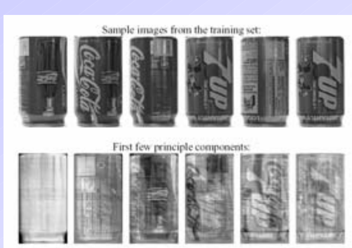
- Eigen-pyramid: basis at multiresolution



Figures from M. J. Black and A. D. Jepson " EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation" ECCV 1996

Eigen-tracking

- Michael J. Black and Allan D. Jepson " EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation"



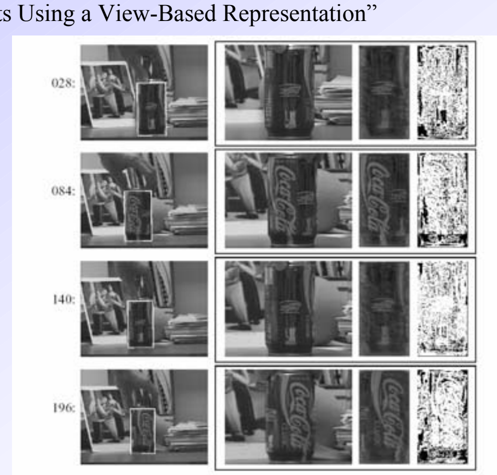


Fig. 8. EigenTracking with translation and divergence over 200 frames. The soda can rotates about its major axis while moving relative to the camera.

Figures from M. J. Black and A. D. Jepson " EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation" ECCV 1996

Eigen-tracking

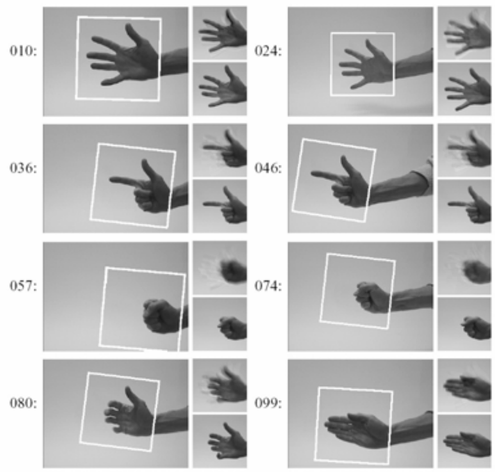
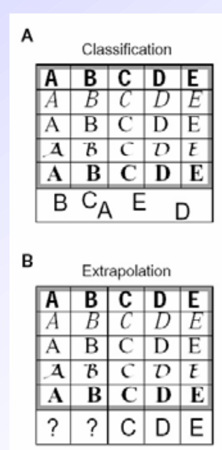


Fig. 10. Tracking and recognizing hand gestures in video.

Figures from M. J. Black and A. D. Jepson " EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation" ECCV 1996

Separating Style and Content

- Objective: Decomposing two factors using linear methods
  - Content: which character
  - Style : which font
- "Bilinear models"
- J. Tenenbaum and W. Freeman "Separating Style and Content with Bilinear Models" Neural computation 2000

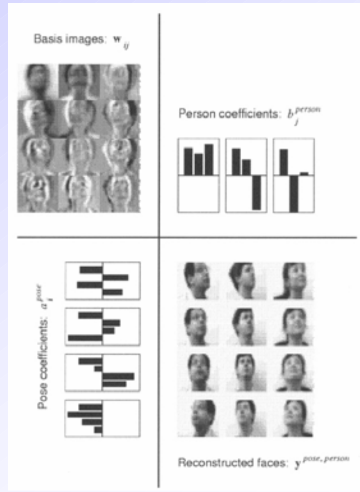


Figures from J. Tenenbaum and W. Freeman "Separating Style and Content with Bilinear Models" Neural computation 2000

Bilinear Models

- Symmetric bilinear model

$$y^{sc} = \sum_{i,j} w_{ij} a_i^s b_j^c$$

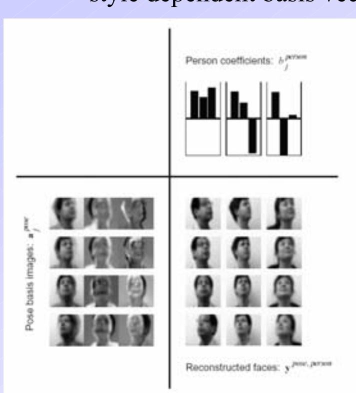


Figures from J. Tenenbaum and W. Freeman "Separating Style and Content with Bilinear Models" Neural computation 2000

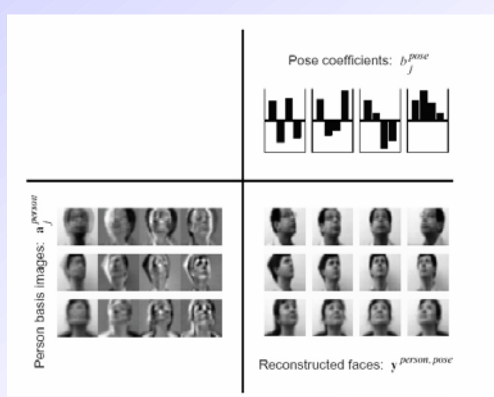
Bilinear models

- Asymmetric bilinear model: use style dependent basis vectors

$$y^{sc} = A^s b^c$$

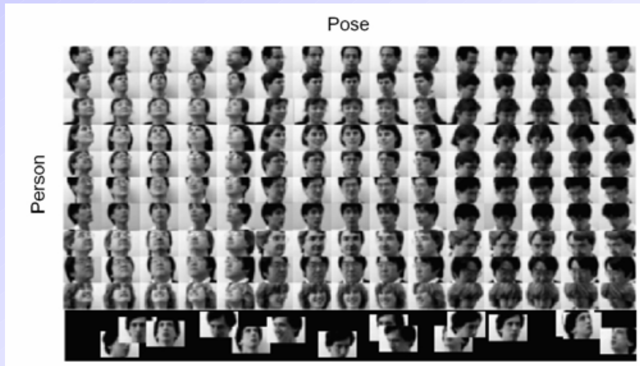


Head pose as style factor
person as content



Person as style factor
pose as content

Figures from J. Tenenbaum and W. Freeman "Separating Style and Content with Bilinear Models" Neural computation 2000



Figures from J. Tenenbaum and W. Freeman "Separating Style and Content with Bilinear Models" Neural computation 2000

Sources

- S. K. Nayar et al 1996 "RealTime 100 Object Recognition System" Technical Report CUCS-019-95, September 1994. Proceedings of ARPA Image Understanding Workshop, San Fransisco, February 1996.
- S. K. Nayar, H. Murase, and S. A. Nene, "Parametric Appearance Representation," in Early Visual Learning, edited by S. K. Nayar and T. Poggio, Oxford University Press, February 1996.
- M. Turk and A. Pentland "Eigenfaces for Recognition" J. Cognitive Neuroscience, vol. 3, pp. 71--86, 1994
- T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. "Active shape models: Their training and application." (1995) CVIU, 61(1):38-59 –
- Many other useful publications and information about Active shape and Active appearance models can be found at T. Cootes we page: http://www.isbe.man.ac.uk/~bim/
- M. J. Black and A. D. Jepson " EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation" ECCV 1996
- Figures from J. Tenenbaum and W. Freeman "Separating Style and Content with Bilinear Models" Neural computation 2000