Recognition Challenges

- Within-class variability
  - Different objects within the class have different shapes or different material characteristics
  - Deformable
  - Articulated
  - Compositional
- Pose variability:
  - 2-D Image transformation (translation, rotation, scale)
  - 3-D Pose Variability (perspective, orthographic projection)
- Lighting
  - Direction (multiple sources & type)
  - Color
  - Shadows
- Occlusion – partial
- Clutter in background -> false positives

Many issues

- Instances vs. categories
- Object representation: 2-D, 3-D, primitives, volumes, points, lines, color, dynamics, function,
- Image features: Intensities, Color, Texture, Shape (2D or 3-D), Motion
- Classification method
- Search strategies
- Automatically learned or taught

Example: Face Detection

- Scan window over image.
- Classify window as either:
  - Face
  - Non-face
Pattern Classification Summary

• Supervised vs. Unsupervised: Do we have labels?
• Supervised
  – Bayesian
    • Plug in classifier
    • Distribution-based
    • Projection Methods (Fisher’s, LDA)
  – Nearest Neighbor
  – Neural Network
  – Support Vector Machine
  – Kernel methods
• Unsupervised
  – Clustering
  – Reinforcement learning

Image as a Feature Vector

• Consider an n-pixel image to be a point in an n-dimensional space, \( x \in \mathbb{R}^n \).
• Each pixel value is a coordinate of \( x \).

Simplest Recognition Scheme

• \( I \) is an image.
• \( c(R_j, I) \) is Euclidean distance.

Nearest Neighbor Classifier

Eigenfaces: linear projection

• An n-pixel image \( x \in \mathbb{R}^n \) can be projected to a low-dimensional feature space \( y \in \mathbb{R}^m \) by \( y = Wx \) where \( W \) is an \( n \) by \( m \) matrix.

Eigenfaces: Principal Component Analysis (PCA)
Assume we have a set of \( n \) feature vectors \( x_i \) (\( i = 1, \ldots, n \)) in \( \mathbb{R}^d \). Write
\[
\mu = \frac{1}{n} \sum_{i=1}^{n} x_i, \\
\Sigma = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu)(x_i - \mu)^T
\]
The unit eigenvectors of \( \Sigma \) — which we write as \( v_1, v_2, \ldots, v_k \), where the order is given by the size of the eigenvalue and \( v_1 \) has the largest eigenvalue — give a set of features with the following properties:
• They are independent.
• Projection onto the basis \( \{v_1, \ldots, v_k\} \) gives the k-dimensional set of linear features that preserves the most variance.

Algorithm 22.5: Principal components analysis identifies a collection of linear features that are independent, and capture as much variance as possible from a dataset.
Some details: Use Singular value decomposition, “trick” described in text to compute basis when \( n \ll d \)
Eigenfaces

- **Modeling**
  1. Given a collection of n labeled training images,
  2. Compute mean image and covariance matrix.
  3. Compute k Eigenvectors (note that these are images) of covariance matrix corresponding to k largest Eigenvalues.
  4. Project the training images to the k-dimensional Eigenspace.

- **Recognition**
  1. Given a test image, project to Eigenspace.
  2. Perform classification to the projected training images.

Eigenfaces: Training Images

Variable Lighting

Reconstruction using Eigenfaces

- Given image on left, project to Eigenspace, then reconstruct an image (right).

Underlying assumptions

- Background is not cluttered (or else only looking at interior of object)
- Lighting in test image is similar to that in training image.
- No occlusion
- Size of training image (window) same as window in test image.
Difficulties with PCA

- Projection may suppress important detail
  - smallest variance directions may not be unimportant
- Method does not take discriminative task into account
  - typically, we wish to compute features that allow good discrimination
  - not the same as largest variance

Illumination Variability

“The variations between the images of the same face due to illumination and viewing direction are almost always larger than image variations due to change in face identity.”

– Moses, Adini, Ullman, ECCV ’94

Fisherfaces: Class specific linear projection


- An n-pixel image $x \in \mathbb{R}^n$ can be projected to a low-dimensional feature space $y \in \mathbb{R}^m$ by
  \[ y = Wx \]
  where $W$ is an $n \times m$ matrix.
- Recognition is performed using nearest neighbor in $\mathbb{R}^m$.
- How do we choose a good $W$?

PCA & Fisher’s Linear Discriminant

- Between-class scatter
  \[ S_b = \sum_{i=1}^{c} |\chi_i - \mu| (\chi_i - \mu)^T \]
- Within-class scatter
  \[ S_w = \sum_{i=1}^{c} |x_i - \mu_i| (x_i - \mu_i)^T \]
- Total scatter
  \[ S_t = \sum_{i=1}^{c} |x_i - \mu| (x_i - \mu)^T = S_b + S_w \]
- How do we choose a good $W$?

Where
- $c$ is the number of classes
- $\mu_i$ is the mean of class $\chi_i$
- $|\chi_i|$ is number of samples of $\chi_i$. 
PCA & Fisher’s Linear Discriminant

- PCA (Eigenfaces)
  \[ W_{PCA} = \arg\max_W \|W^T S W \| \]
  Maximizes projected total scatter

- Fisher’s Linear Discriminant
  \[ W_{FLD} = \arg\max_W \frac{W^T S_W W}{W^T S_B W} \]
  Maximizes ratio of projected between-class to projected within-class scatter

Fisherfaces

- Since \( S_W \) is rank N-c, project training set to subspace spanned by first N-c principal components of the training set.
- Apply FLD to N-c dimensional subspace yielding c-1 dimensional feature space.

- Fisher’s Linear Discriminant projects away the within-class variation (lighting, expressions) found in training set.
- Fisher’s Linear Discriminant preserves the separability of the classes.

Harvard Face Database

- 10 individuals
- 66 images per person
- Train on 6 images at 15°
- Test on remaining images

Recognition Results: Lighting Extrapolation

- Error Rate
- Correlation
- Eigenfaces
- Eigenfaces (w/o 1st 3)
- Fisherfaces

Variability: Camera position
  Illumination
  Internal parameters
  Within-class variations
Appearance-based vision for robot control

[ Nayar, Nene, Murase 1994 ]

Limitations of these approaches

- Object must be segmented from background (How would one do this in non-trivial situations?)
- Occlusion?
- The variability (dimension) in images is large, so is sampling feasible?
- How can one generalize to classes of objects?

Appearance-Based Vision: Lessons

Strengths
- Posing the recognition metric in the image space rather than a derived representation is more powerful than expected.
- Modeling objects from many images is not unreasonable given hardware developments.
- The data (images) may provide a better representations than abstractions for many tasks.

Weaknesses
- Segmentation or object detection is still an issue.
- To train the method, objects have to be observed under a wide range of conditions (e.g. pose, lighting, shape deformation).
- Limited power to extrapolate or generalize (abstract) to novel conditions.