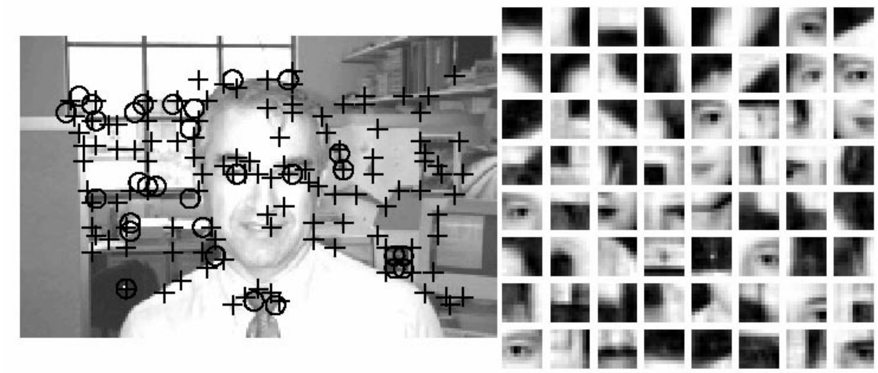
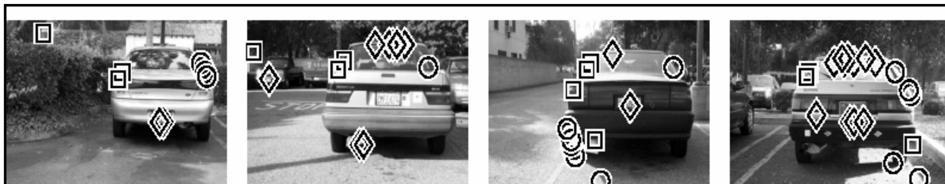
Unsupervised Learning of Models for Recognition

Weber/Perona

Automatically figure out which objects are present
in those images………………………………….but not in these



Large number of interest points from training images → Clustering → Small number of representative parts (cluster centers)



= A   = B   = C

Data Representation

x_{ij} Position of *j*-th occurrence of part *i*

$$X^{o} = \begin{pmatrix} x_{11} & \cdots & x_{1N_1} \\ \vdots & \ddots & \vdots \\ x_{T1} & \cdots & x_{TN_2} \end{pmatrix}$$

Observables

Number of occurrences

Hidden data (unobservable)

- Vector $\mathbf{h}$ of length = number of parts F on the object of interest

$$\mathbf{h}_i = \begin{cases} j & \text{If part } i \text{ at position } j \text{ is on the foreground (the object of interest)} \\ 0 & \text{If part } i \text{ is undetected or is not detected on the object (accounts for occlusion and missed detections)} \end{cases}$$

- $\mathbf{b}$ binary version of $\mathbf{h}$

$$\mathbf{b}_i = \begin{cases} 1 & \text{If part } i \text{ occurs in the foreground} \\ 0 & \text{If part } i \text{ is in a background part} \end{cases}$$

- $\mathbf{n}$ number of background occurrences for each part
- $\mathbf{n}$ and $\mathbf{b}$ are completely defined by $\mathbf{h}$
- $\mathbf{x}^\mathbf{m}$ missing parts from the object

$$P(\mathbf{X}^o, \mathbf{x}^m, \mathbf{h}) = P(\mathbf{X}^o, \mathbf{x}^m, \mathbf{h}, \mathbf{n}, \mathbf{b})$$

Given the number of occurrences on foreground/background, what are the possible labelings?

$$= P(\mathbf{X}^o, \mathbf{x}^m | \mathbf{h}, \mathbf{n}) P(\mathbf{h} | \mathbf{b}, \mathbf{n}) P(\mathbf{n}) P(\mathbf{b})$$

Given a configuration foreground/background, what are the likely configurations of parts locations?

Distributions of parts on background and foreground independently of their positions

$$\boxed{P(\mathbf{X}^o, \mathbf{x}^m \mid \mathbf{h}, \mathbf{n})} P(\mathbf{h} \mid \mathbf{b}, \mathbf{n}) P(\mathbf{n}) P(\mathbf{b})$$

Coordinates of all the foreground parts (including the missing parts)

Coordinates of all the background parts

$$P_{fg}\left(\boxed{\mathbf{x}^o, \mathbf{x}^m}\right) P_{bg}\left(\boxed{\mathbf{x}_{bg}}\right)$$

Gaussian distribution (μ,Σ)

Uniform distribution

$$P(\mathbf{X}^o, \mathbf{x}^m \mid \mathbf{h}, \mathbf{n}) \boxed{P(\mathbf{h} \mid \mathbf{b}, \mathbf{n})} P(\mathbf{n}) P(\mathbf{b})$$

- Uniform distribution over **h**'s consistent with **b** and **n** No parameter to estimate

$$P(\mathbf{X}^o, \mathbf{x}^m \mid \mathbf{h}, \mathbf{n}) P(\mathbf{h} \mid \mathbf{b}, \mathbf{n}) \boxed{P(\mathbf{n})} P(\mathbf{b})$$

- Poisson distribution parameter M_t = average number of detection of part type t in the background

$$P(\mathbf{X}^o, \mathbf{x}^m | \mathbf{h}, \mathbf{n}) P(\mathbf{h} | \mathbf{b}, \mathbf{n}) P(\mathbf{n}) \boxed{P(\mathbf{b})}$$

- Table of probabilities of co-occurrences of object parts.
- P([1 0 0 1 1]) = probability that parts 1, 4, and 5 are detected on the foreground object and that 2,3 are not
- Minor tiny issue: Must use a table of size 2^F

Model Summary

- Need to estimate from training images:
  - μ and Σ $\left(P(\mathbf{X}^o, \mathbf{x}^m | \mathbf{h}, \mathbf{n})\right)$
  - $\mathbf{M}$ ($P(\mathbf{n})$)
  - $P(\mathbf{b})$

$$\theta = (\mu, \Sigma, P(\mathbf{b}), \mathbf{M})$$

Parameter Estimation

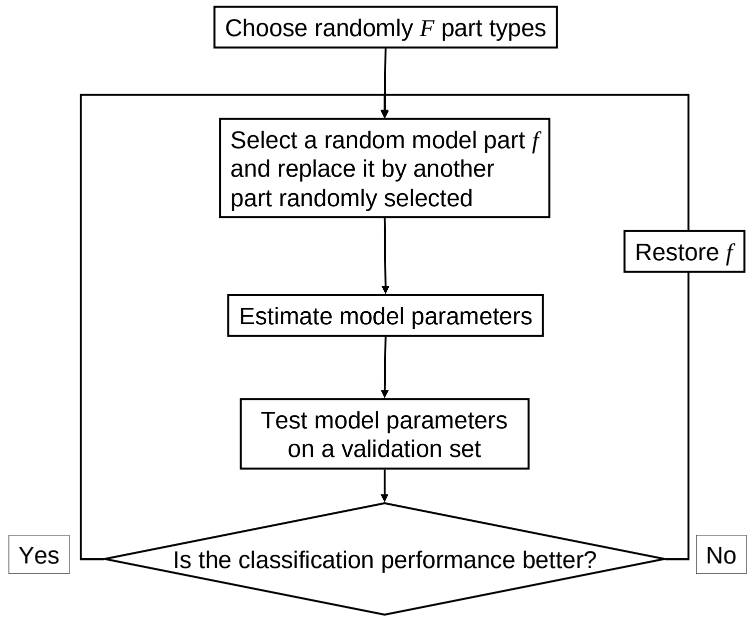
- Given training images i = 1,..,I
- M-step: Find θ that maximizes the data likelihood, given the distribution of hidden variables (**h**)

- E-step: Estimate the posterior distribution of hidden variables (**h**) using the current value of θ: $P(\mathbf{h}_i, \mathbf{x}_i^m \mid X_i^o, \theta)$

Parts Selection

- Problem: We have a collection of N part types from the training data, we want to select a small number F that is representative of the object
- F must be small because of $P(\mathbf{b})$
- Tiny problem: N^F possible selections!!!

Parts Selection



Classification

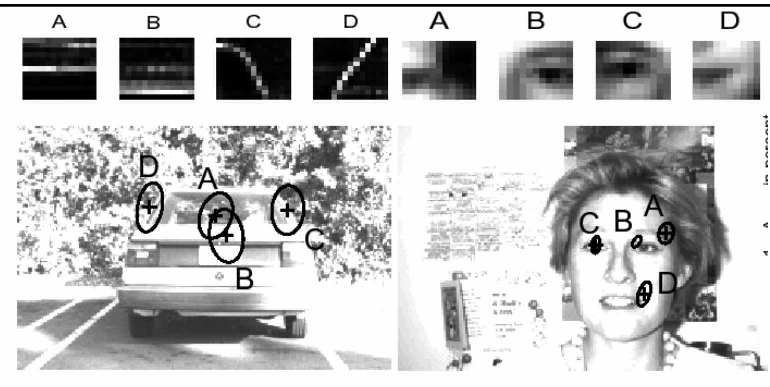
Given input image
- Match parts from training image with parts from training data X^o
- Decide between hypotheses:
  - C_1: The object is present in the image
  - C_2: The object is not present

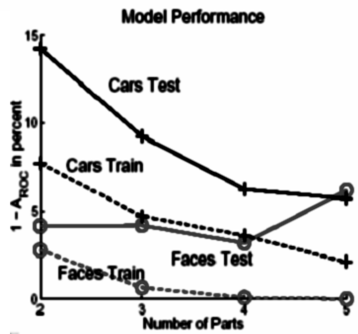
Sum over all possible configurations in which the parts may appear if the object is present

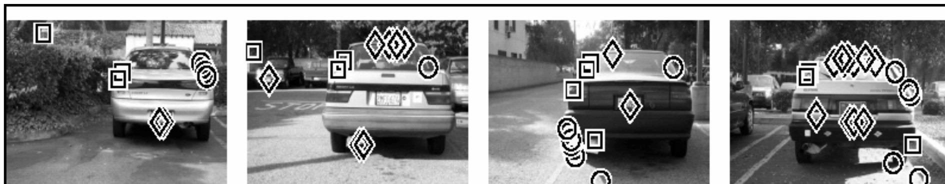
Evaluate and threshold:

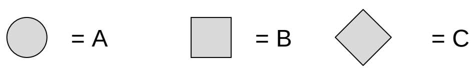
$$\frac{P\left(C_1 \middle| X^o\right)}{P\left(C_2 \middle| X^o\right)} \sim \frac{\sum_{\mathbf{h}} P\left(X^o, \mathbf{h} \middle| C_1\right)}{P\left(X^o, \mathbf{h}_o \middle| C_o\right)}$$

Null hypothesis all the parts are in the background


A
B
C
D
A
B
C
D
D
A
B
C
C
B
A
D


Model Performance
Cars Test
Cars Train
Faces Test
Faces Train
1 - A_ROC in percent
Number of Parts
15
10
5
0
2
3
4
5




= A
= B
= C


A
B
C

Generalization: Mixture of models

Probability of observing component ω of the model

$$P(X^o,\mathbf{h}) = \sum_{\omega} P(X^o,\mathbf{h}|\omega)P(\omega)$$

Distribution for component ω of the model

Thoughts

- Allows discovery of models from unlabelled data
- Based on generative representation of the relations between parts
- Major combinatorial issue potentially