Lecture 17: Parts-based models and context
Announcements

• Project 3: Eigenfaces
  – due Wednesday, November 11 at 11:59pm
  – solo project
Object → Bag of ‘words’
What about spatial info?
Problem with bag-of-words

- All have equal probability for bag-of-words methods
- Location information is important
Model: Parts and Structure
Deformable objects

Images from D. Ramanan’s dataset
Deformable objects

Images from Caltech-256
Part-based representation

• Objects are decomposed into parts and spatial relations among parts

Fischler and Elschlager ‘73
Pictorial structures

• Two components:
  – Appearance model
    • How much does a given window look like a given part?
  – Spatial model
    • How well do the parts match the expected shape?
Formal Definition of Model

- Set of parts $V = \{v_1, \ldots, v_n\}$
Pictorial Structure

• Matching = Local part evidence + Global constraint

\[ L^* = \arg \min_L \left( \sum_{i=1}^{n} m_i(l_i) + \sum_{(v_i,v_j) \in E} d_{ij}(l_i, l_j) \right) \]

• \( m_i(l_i) \): matching cost for part \( I \)
• \( d_{ij}(l_i, l_j) \): deformable cost for connected pairs of parts
• \( (v_i,v_j) \): connection between part \( i \) and \( j \)
Flexible Template Algorithms

- Difficulty depends on structure of graph
  - Which parts connected and form of constraint
Part-based representation

- Tree model \(\Rightarrow\) Efficient inference by dynamic programming
Appearance model

• Each part has an associated appearance model
  – E.g., a reference patch, gradient histogram, etc.
Spatial model

• Each edge represents a spring with a certain relative offset, covariance
Matching on tree structure

\[ E(L) = \sum_{i=1}^{n} m_i(l_i) + \sum_{(v_i,v_j) \in E} d_{ij}(l_i,l_j) \]

- For each \( l_1 \), find best \( l_2 \):

\[ \text{Best}_2(l_1) = \min_{l_2} \left[ m_2(l_2) + d_{12}(l_1,l_2) \right] \]

- Remove \( v_2 \), and repeat with smaller tree, until only a single part

- Complexity: \( O(nk^2) \): \( n \) parts, \( k \) locations per part
Putting it all together
Sample result on matching human
Sample result on matching human
Matching results
Learning the model parameters

• Easiest approach: supervised learning
  – Someone chooses the number and meaning of the parts, labels them in a bunch of training examples
  – Use this to learn the appearance and spatial models

• A lot of work has been done on unsupervised learning of these models
Some learned object models
Part-based representation

• K-fans model (D.Crandall, et.al., 2005)

Figure 1. Some k-fans on 6 nodes. The reference nodes are shown in black while the regular nodes are shown in gray.
How much does shape help?

- Crandall, Felzenszwalb, Huttenlocher CVPR’05
- Shape variance decreases with increasing model complexity
- Do get some benefit from shape
3-minute break
Context: thinking outside the (bounding) box

© Oliva & Torralba

Slides courtesy Alyosha Efros
Eye of the Beholder

Claude Monet

Gare St. Lazare

Paris, 1877
Eye of the Beholder

where did it go?
Seeing less than you think…
Seeing less than you think...
What the Detector Sees
What the Detector Does

True Detection

Missed

False Detections

Local Detector: [Dalal-Triggs 2005]
If we have 1000 categories (detectors), and each detector produces 1 FP every 10 images, we will have 100 false alarms per image… pretty much garbage…

Slide by Antonio Torralba
Context to the rescue!

We know there is a keyboard present in this scene even if we cannot see it clearly.

We know there is no keyboard present in this scene

... even if there is one indeed.

Slide by Antonio Torralba
When is context helpful?

Local features

Contextual features

Distance

Information

Slide by Antonio Torralba
Is it just for small / blurry things?

A  B  C
Is it just for small / blurry things?
Is it just for small / blurry things?
Context is hard to fight!

Thanks to Paul Viola for showing me these
more “Look-alikes”
Don’t even need to see the object
Don’t even need to see the object

Chance ~ 1/30000
But object can say a lot about the scene

The influence of an object extends beyond its physical boundaries