part 6: summary
Recognition System

- **Representation**: use a set of basic features
- **Learning**: learn representation of objects from training images (sometimes of minimal size or larger)
- **Recognition**: decide whether the object exists in the image
  - Detection
  - Localization
  - Segmentation
Representation

First challenge: obtain informative representation

- Global representation - all parts of the image or window impact the description
  ⇒ sensitive to occlusion, clutter, viewpoint

- Local features representation - describe component regions separately

Eigen faces [Turk & Pentland 91]
Shape context [Belongie 02]
HOG [Dalal & Triggs, 05]
Integral images [Viola & Jones, 04]

SIFT [Lowe 99]
Maximally Stable Extremal Regions [Matas 02]
Bag of words
Constellation model [Fergus et al 03]
Learning:

- When number of features is not prohibitive (e.g., quantized color values)
  - Vector (ordered list) of all features
  - Histogram (distribution) of feature values (bag of words)

- Group of possible features is very large
  - Set of features in object
  - Set of object features, related to each other by a Graphical model (constellation models)

- This representation is learned, or computed, from a set of training images
Recognition

- Compute global features
  ⇒ sliding window

- Compute local features in the image
  ⇒ do one of:
    - Intersect set of image features with set of object features
    - Compute probability for image features based on object’s graphical model
    - Check whether the image is a permissible transformation of the object
Extensions: paper presentations

1. Combining steps: representation and learning
2. Sliding window: efficient sub-window search

Better features:
3. Hierarchy of parts
4. Cortex-like features
5. Matching: learning distance function
6. Coupled detection and trajectory estimation

Unsupervised exploration with pLSA and LDA
7. category discovery
8. Scene classification
Extensions: paper presentations

9. Representation: discriminative or invariant?
10. Dataset issues
11. Discriminative global part-based model
12. Working with video: learning human actions
13. Multi-View recognition
14. Exploring classifiers: random forests and ferns
15. Using context and scene layout
16. Knowledge transfer between classes
3. Better features: part hierarchy

- $L_2, L_3$ (non-specific)
- $L_4$ (category specific)
$L_5$
4. Better features: cortex-like features

(SMF1) (SMF2)
5. Learning distance function

Categorization scenario
6. Coupled detection and trajectory estimation

Goal: combine object tracking with detection

Approach:
Related work: integrating segmentation & recognition

Kumar et al. CVPR 2005

Borenstein & Ullman, ECCV 2002

Tu, Chen, Yuille, Zhu, ICCV 2003

Kannan, Winn, & Rother, NIPS 2006
7. Unsupervised category discovery

Topic models for images

Probabilistic Latent Semantic Analysis (pLSA)

Latent Dirichlet Allocation (LDA)

Sivic et al. ICCV 2005, Fei-Fei et al. ICCV 2005

Figure credit: Fei-Fei Li
8. Scene classification

Feature Extraction

Training

Bag of words

Test

Visual Vocabulary

$w_1, w_2, \ldots, w_p, w_q, w_r, \ldots, w_k$

Training Images

Test Image

Bag of words

Classification

Learning

pLSA

$P(w|d_{training})$

$P(w|z)$

$P(z|d_{training})$

$pLSA (fixed P(w|z))$

$P(w|d_{test})$

$P(w|z)$

$P(z|d_{test})$

Similarity & KNN classification

K most similar images
9. Representation: discriminative or invariant?

Task Specific Trade-Off:

Don’t want rotation invariance

Don’t want size invariance

What is the right amount of invariance?

Solution: Sparse Multiple Kernel Learning Classification Formulation
10. Dataset issues

Example: Caltech-101

A dataset that has been about mastered...

Images from the Caltech-101: 101-way multi-class classification problem

Slide credit: Grauman & Leibe
Example: Caltech256

Images from the Caltech-256:
256 multi-class recognition problem

Slide credit: Grauman & Leibe
Example: Pascal Visual Object Classes Challenge

Pascal VOC 2007:
Binary detection problems

http://pascallin.ecs.soton.ac.uk/challenges/VOC/

Slide credit: Grauman & Leibe
Example: LabelMe

http://labelme.csail.mit.edu/

Slide credit: Grauman & Leibe
11. Discriminative global part-based model
12. learning human actions

Learning the Action Models: pLSA

walking, running, jogging...

walking
running
jogging

boxing
hand clapping
hand waving
13. Multi-Aspect Recognition

Detectors for different viewpoints ⇒ How can this be improved?
14. Exploring classifiers: random forests and ferns
15. understanding scene layout
16. Knowledge transfer between classes

Performance comparison

Sample ROC of different learning models

Shape Model (Training # = 1)

Appearance Model (Training # = 1)
Summary

- Object recognition is a hard problem
- A lot of progress has been made in the last few years when working with realistic scenes
- There is a lot more work to be done…