outline

• Discriminative vs. Generative Classifiers
• Image representation and recognition models
  – Bag of Words Model
  – Part-based Model
    • Constellation Model
    • Pictorial Structures Model
  – Spatial Pyramid Matching (SPM)
  – ObjectBank
Discriminative vs Generative Classifiers

Training classifiers involves estimating \( f: X \rightarrow Y \), or \( P(Y|X) \)

- **Discriminative classifiers** (e.g. logistic regression, SVM):
  - We want to model \( P(Y|X) \)
  - Assume some functional form for \( P(Y|X) \)
  - Estimate parameters of \( P(Y|X) \) directly from training data

- **Generative classifiers** (e.g. naïve bayes):
  - We want to model \( P(X, Y) \)
  - Assume some functional form for \( P(X|Y) \), \( P(X) \)
  - Estimate parameters of \( P(X|Y) \), \( P(X) \) directly from training data
  - Use Bayes rule to calculate \( P(Y|X=x_i) \)
Discriminative vs Generative Classifiers

• Advantages of discriminative classifiers:
  – Typically faster at making predictions
  – Tend to have better performance
  – Direct modeling of what we want to optimize

• Advantage of generative classifiers:
  – Can handle missing/partially labeled data
  – A new class (Y+1) can be added incrementally without training the complete model
  – Can generate samples from the training distribution

[Ulusoy & Bishop, 2005]
Image Representation

Bag of Words

Weakly Spatial Models

- Spatial Pyramid Matching
- Object-Bank

Part-based

- Constellation Model
- Pictorial Structure

Spatial Specificity of Parts

No spatial info. Very strong but sparse

Fei-Fei Li
Image Representation

Bag of Words

- Weakly Spatial Models
  - Constellation Model
  - Pictorial Structure
- Spatial Models
  - Spatial Pyramid Matching
  - Object-Bank
- Part-based
  - Constellation Model
  - Pictorial Structure

Spatial Specificity of Parts

No spatial info. Very strong but sparse

Fei-Fei Li
TA Section - 4
6
22-Nov-11
Bag-of-words Representation

Legend
- Training local feature
- Cluster 1
- Cluster 2
- Cluster 3
- Test local feature

Local features (e.g. SIFT features)

Clustering (e.g. K-means)

Dictionary or Code book

Using minimum Euclidean distance!

Assign feature to a dictionary word

Bag-of-words!

Normalization (Why?)

Histogram Assignment

Count

Feature generation

Fei-Fei Li

TA Section - 4

7

22-Nov-11
Image Representation

- Bag of Words
- Weakly Spatial Models
- Spatial Pyramid Matching
- Object Bank

Part-based

- Constellation Model
- Pictorial Structure

Spatial Specificity of Parts

No spatial info.  Very strong but sparse
Image Representation

- Bag of Words
- Weakly Spatial Models
  - Spatial Pyramid Matching
  - Object Bank

- Part-based
  - Constellation Model
  - Pictorial Structure

Spatial Specificity of Parts

No spatial info.

Very strong but sparse
Generative probabilistic model (2)

Foreground model

Gaussian shape pdf

Gaussian part appearance pdf

Gaussian relative scale pdf

Prob. of detection

0.8 0.75 0.9

Clutter model

Uniform shape pdf

Gaussian background appearance pdf

Uniform relative scale pdf

Poission pdf on # detections
Image Representation

Bag of Words

Weakly Spatial Models

Part-based

- Constellation Model
- Pictorial Structure

No spatial info.

Spatial Specificity of Parts

Very strong but sparse
Pictorial Structures

• Basic idea:
  We would like to represent an object by
  – a collection of parts
  – arranged in a deformable configuration
Pictorial Structures

• Local model of appearance with non-local geometric or spatial constraints
• Simultaneous use of appearance and spatial information
  – Simple part models alone are not discriminative
• The model needs to solve the tasks:
  – determine whether an object is visible in an image
  – determine where an object is in the image
Pictorial Structures

- Model is represented as an undirected graph structure $G = (V, E)$, where $V$ are the vertices and $E$ are the edges.

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

- Matching score of individual parts
- Sum over all parts
- Deformation score of connected parts
Image Representation

Weakly Spatial Models

- Bag of Words
- Constellation Model
- Pictorial Structure

- Spatial Pyramid Matching

Spatial Specificity of Parts

No spatial info. Very strong but sparse
Start with Pyramid Matching Kernel for BoW Models

\[ X = \{ \bar{x}_1, \ldots, \bar{x}_m \} \quad \text{Sets of features} \quad Y = \{ \bar{y}_1, \ldots, \bar{y}_n \} \]

[Grauman & Darrell, 2005]
Pyramid Matching Kernel

• How do we build a **discriminative classifier** using the set representation?

• Kernel-based methods (e.g. SVM) are appealing for efficiency and generalization power.

• But what is an appropriate kernel?
  – Each instance is an unordered set of vectors
  – Varying number of vectors per instance
Pyramid Matching Kernel

• We can compare sets by computing a **partial matching** between their features

\[ K_\Delta = \sum_{i=0}^{L} w_i N_i \]

- Number of newly matched pairs at level \( i \)
- Measure of difficulty of a match at level \( i \)
- Approximate partial match similarity
Pyramid Matching Kernel (Example)

Level 0

\[ N_0 = 2 \]
\[ w_0 = 1 \]

\[ \mathcal{I}_0 = 2 \]
Pyramid Matching Kernel (Example)

Level 1

\[ N_1 = 4 - 2 = 2 \]
\[ w_1 = \frac{1}{2} \]

\[ H_1(X) \]
\[ H_1(Y) \]
\[ J_1 = 4 \]
Pyramid Matching Kernel

optimal partial matching between sets of features

\[
K_\Delta (\Psi(X), \Psi(Y)) = \frac{1}{2^i} \left( I(H_i(X), H_i(Y)) - I(H_{i-1}(X), H_{i-1}(Y)) \right)
\]

difficulty of a match at level \(i\)

number of new matches at level \(i\)

Fei-Fei Li
Spatial Pyramid Matching

• Pyramid Match Kernel (Grauman & Darrell)
  \( \text{Pyramid in feature space, ignore location} \)

• Spatial Pyramid (Lazebnik et al)
  \( \text{Pyramid in image space, quantize features} \)

Features:

- Weak (edge orientations)
- Strong (SIFT)
Spatial Pyramid Matching

level 0

level 1

level 2

Fei-Fei Li
Spatial Pyramid Matching

Feature histograms:

Level 3

Level 2

Level 1

Level 0

Total weight (value of pyramid match kernel): \( I_3 + \frac{1}{2}(I_2 - I_3) + \frac{1}{4}(I_1 - I_2) + \frac{1}{8}(I_0 - I_1) \)
Image Representation

Bag of Words

Weakly Spatial Models

- Constellation Model
- Pictorial Structure

Part-based

Spatial Specificity of Parts

No spatial info.

- Spatial Pyramid Matching

• Object-Bank

Very strong but sparse

Fei-Fei Li
Event: “Sailing”

Object Bank

High Level Objects based
Sailboat, water, sky, tree, ...

Image representation

Semantic Gap

High level tasks

Low level feature
HoG, Gist, SIFT, Color, Texture, Bag of Words (BoW), Spatial Pyramid (SPM)

Li et al. 2010

Fei-Fei Li
Object Bank representation

Object Filters  Spatial pooling

Sailboat response

Li et al. 2010
Object Bank representation

Object Filters  Spatial pooling

Sailboat response

Object size variance

- Small
- Median
- Large

Li et al. 2010
Object Bank representation

Object Filters Spatial pooling

Sailboat response

Bear response

Water response

Implementation details

• ~ 200 object detectors
• Felzenswalb et al. 2008
• Hoeim et al. 2005
• 3-level spatial pyramid
• for each grid: max of each object
Object Bank representation

Object Filters  Spatial pooling

Sailboat response

Bear response

Water response

E.g., In our setting: $12 \times 21 \times 177 = 44604$

$N: nGridsperScale,
M: nScale,
O: nObject$

$M \times N \times O = \text{thousands}$

Li et al. 2010
A word about Q2 in PS4

• We’d like you to understand the differences between BoW, SPM, and ObjectBank
• We’d like you to use what you’ve learned so far, be creative, and come up with interesting ways of encoding image information for an image recognition task
• Extra credits are given especially to innovation and good performances