Beyond bags of features: Part-based models

Many slides adapted from Fei-Fei Li, Rob Fergus, and Antonio Torralba
Implicit shape models

• Visual codebook is used to index votes for object position

training image annotated with object localization info

B. Leibe, A. Leonardis, and B. Schiele, Combined Object Categorization and Segmentation with an Implicit Shape Model, ECCV Workshop on Statistical Learning in Computer Vision 2004
Implicit shape models

- Visual codebook is used to index votes for object position

test image

B. Leibe, A. Leonardis, and B. Schiele, Combined Object Categorization and Segmentation with an Implicit Shape Model, ECCV Workshop on Statistical Learning in Computer Vision 2004
Implicit shape models: Training

1. Build codebook of patches around extracted interest points using clustering
Implicit shape models: Training

1. Build codebook of patches around extracted interest points using clustering
2. Map the patch around each interest point to closest codebook entry
Implicit shape models: Training

1. Build codebook of patches around extracted interest points using clustering
2. Map the patch around each interest point to closest codebook entry
3. For each codebook entry, store all positions it was found, relative to object center
Implicit shape models: Testing

1. Given test image, extract patches, match to codebook entry
2. Cast votes for possible positions of object center
3. Search for maxima in voting space
4. Extract weighted segmentation mask based on stored masks for the codebook occurrences
Example: Results on Cows

Original image

Source: B. Leibe
Example: Results on Cows

Source: B. Leibe
Example: Results on Cows

Matched patches

Source: B. Leibe
Example: Results on Cows

Probabilistic votes

Source: B. Leibe
Example: Results on Cows

Hypothesis 1

Source: B. Leibe
Example: Results on Cows

Hypothesis 2

Source: B. Leibe
Example: Results on Cows

Hypothesis 3

Source: B. Leibe
Additional examples

Generative part-based models

What is a generative model?

- Model the probability of an image given a class

\[ p(\text{image} \mid \text{zebra}) \] vs. \[ p(\text{image} \mid \text{no zebra}) \]
Making a decision

- **Maximum a posteriori (MAP) decision**: assign the image to the class that gets the highest *posterior probability* $P(\text{class} | \text{image})$. 

![Zebra Image](image.jpg)
Making a decision

- **Maximum a posteriori (MAP) decision:** assign the image to the class that gets the highest *posterior probability* $P(\text{class} \mid \text{image})$.

- **Bayes rule:**

  $$p(\text{class} \mid \text{image}) = \frac{p(\text{image} \mid \text{class})p(\text{class})}{p(\text{image})}$$
Generative vs. discriminative models

\[ p(\text{class} | \text{image}) \propto p(\text{image} | \text{class}) p(\text{class}) \]

- **Generative methods**: model likelihood and prior
- **Discriminative methods**: model posterior
Generative vs. discriminative models

Generative

\[ p(x|C_1) \quad p(x|C_2) \]

Class densities

Discriminative

\[ p(C_1|x) \quad p(C_2|x) \]

Posterior probabilities
The Naïve Bayes model

- Generative model for a bag of features
- Assume that each feature $f_i$ is conditionally independent given the class $c$:

$$p(image \mid c) = p(f_1, \ldots, f_N \mid c) = \prod_{i=1}^{N} p(f_i \mid c)$$
Generative part-based model

\[ P(\text{image} \mid \text{object}) = P(\text{appearance, shape} \mid \text{object}) \]
Generative part-based model

\[ P(\text{image} \mid \text{object}) = P(\text{appearance}, \text{shape} \mid \text{object}) = \max_h P(\text{appearance}, \text{shape}, h \mid \text{object}) \]

\( h \): assignment of features to parts
Generative part-based model

\[ P(\text{image} \mid \text{object}) = P(\text{appearance}, \text{shape} \mid \text{object}) \]
\[ = \max_h P(\text{appearance}, \text{shape}, h \mid \text{object}) \]
\[ = \max_h P(\text{appearance} \mid h, \text{object}) p(\text{shape} \mid h, \text{object}) p(h \mid \text{object}) \]

\( h \): assignment of features to parts
Generative part-based model

\[ P(\text{image} \mid \text{object}) = P(\text{appearance}, \text{shape} \mid \text{object}) \]
\[ = \max_h P(\text{appearance} \mid h, \text{object}) p(\text{shape} \mid h, \text{object}) p(h \mid \text{object}) \]
Generative part-based model

\[
P(\text{image} \mid \text{object}) = P(\text{appearance, shape} \mid \text{object})
= \max_h P(\text{appearance} \mid h, \text{object}) p(\text{shape} \mid h, \text{object}) p(h \mid \text{object})
\]
Results: Faces

Face shape model

Recognition results

Patch appearance model
Results: Motorbikes and airplanes
Pictorial structures

• Set of parts (oriented rectangles) connected by edges
• Recognition problem: find the most probable part layout $l_1, \ldots, l_n$ in the image

P. Felzenszwalb and D. Huttenlocher, *Pictorial Structures for Object Recognition*, IJCV 61(1), 2005
Pictorial structures

- **MAP formulation**: maximize posterior

\[
P(l_1, \ldots, l_n \mid \text{Im}) \propto P(\text{Im} \mid l_1, \ldots, l_n)P(l_1, \ldots, l_n) = \prod_i P(\text{Im}(l_i)) \prod_{i,j \in E} P(l_i \mid l_j)
\]

  Appearance  Geometry

- **Energy-based formulation**: minimize minus the log of probability:

\[
E(l_1, \ldots, l_n) = \sum_i m_i(l_i) + \sum_{i,j} d_{ij}(l_i, l_j)
\]

  Matching cost  Deformation cost
Pictorial structures: Complexity

\[ \arg \min_{l_1, \ldots, l_n} \sum_i m_i(l_i) + \sum_{i,j} d_{ij}(l_i, l_j) \]

• Suppose there are \( n \) parts, and each has \( h \) possible positions in the image.

• What is the complexity of finding the optimal layout?

• Brute force search: \( O(h^n) \)

• Tree-structured model: \( O(h^2n) \)

• Deformation cost is quadratic: \( O(hn) \)