Computer Vision - Lecture 15
Part-based Models for Object Categorization
07.01.2016

Bastian Leibe
RWTH Aachen
leibe@vision.rwth-aachen.de

Course Outline
- Image Processing Basics
- Segmentation & Grouping
- Object Recognition
- Object Categorization I
  - Sliding Window based Object Detection
- Local Features & Matching
  - Local Features: Detection and Description
  - Recognition with Local Features
  - Indexing & Visual Vocabularies
- Object Categorization II
  - Bag-of-Words Approaches & Part-based Approaches
  - Deep Learning Methods
- 3D Reconstruction

Topics of This Lecture
- Recap: Specific Object Recognition with Local Features
  - Matching & Indexing
  - Geometric Verification
- Part-Based Models for Object Categorization
  - Structure representations
  - Different connectivity structures
- Bag-of-Words Model
  - Use for image classification
- Implicit Shape Model
  - Generalized Hough Transform for object category detection
- Deformable Part-based Model
  - Discriminative part-based detection

Recap: Recognition with Local Features
- Image content is transformed into local features that are invariant to translation, rotation, and scale
- Goal: Verify if they belong to a consistent configuration

Recap: Indexing features
- Detect or sample features
- List of positions, scales, orientations
- Describe features
- Associated list of d-dimensional descriptors
- Index each one into pool of descriptors from previously seen images
- Match to quantized descriptors (visual words)

⇒ Shortlist of possibly matching images + feature correspondences

Extension: tf-idf Weighting
- Term frequency - inverse document frequency
  - Describe frame by frequency of each word within it, downweight words that appear often in the database
  - (Standard weighting for text retrieval)

Slide credit: Kristen Grauman

Slide credit: David Lowe
Recap: Fast Indexing with Vocabulary Trees

- Recognition

Geometric verification

Recap: Geometric Verification by Alignment

- Assumption
  - Known object, rigid transformation compared to model image
  - If we can find evidence for such a transformation, we have recognized the object.
- You learned methods for
  - Fitting an affine transformation from ≥ 3 correspondences
  - Fitting a homography from ≥ 4 correspondences
    - Affine: solve a system
    - Homography: solve a system
    \[ At = b \]
    \[ Ah = 0 \]
- Correspondences may be noisy and may contain outliers
  - Need to use robust methods that can filter out outliers
  - Use RANSAC or the Generalized Hough Transform

Topics of This Lecture

- Recap: Specific Object Recognition with Local Features
- Part-Based Models for Object Categorization
  - Structure representations
  - Different connectivity structures
- Bag-of-Words Model
  - Use for image classification
- Implicit Shape Model
  - Generalized Hough Transform for object category detection
- Deformable Part-based Model
  - Discriminative part-based detection

Recognition of Object Categories

- We no longer have exact correspondences...
- On a local level, we can still detect similar parts.
- Represent objects by their parts
  - Bag-of-features
- How can we improve on this?
  - Encode structure

Part-Based Models

- Fischler & Elschlager 1973
- Model has two components
  - parts (2D image fragments)
  - structure (configuration of parts)

Different Connectivity Structures

- Bag of visual words
  - Csurka et al. '04
  - Vasconcelos et al. '00
- Constellation
  - Fergus et al. '03
  - Fei-Fei et al. '03
- Star shape
  - Leibe et al. '04, '08
  - Crandall et al. '05
  - Fergus et al. '05
- k-fan (k = 2)
  - Crandall et al. '05
- Hierarchy
  - Bouchard & Triggs '05
- Sparse Flexible model
  - Carneiro & Lowe '06
Topics of This Lecture
- Recap: Specific Object Recognition with Local Features
- Part-Based Models for Object Categorization
  - Structure representations
  - Different connectivity structures
- Bag-of-Words Model
  - Use for image classification
- Implicit Shape Model
- Generalized Hough Transform for object category detection
- Deformable Part-based Model
- Discriminative part-based detection

Recap: Advantage of BoW Histograms
- Bag of words representations make it possible to describe the unordered point set with a single vector (of fixed dimension across image examples).
  - Provides easy way to use distribution of feature types with various learning algorithms requiring vector input.
Limitations of BoW Representations

- The bag of words removes spatial layout.
- This is both a strength and a weakness.
- Why a strength?
- Why a weakness?

Spatial Pyramid Representation

- Representation in-between orderless BoW and global appearance

Summary: Bag-of-Words

- **Pros:**
  - Flexible to geometry / deformations / viewpoint
  - Compact summary of image content
  - Provides vector representation for sets
  - Empirically good recognition results in practice

- **Cons:**
  - Basic model ignores geometry - must verify afterwards, or encode via features.
  - Background and foreground mixed when bag covers whole image
  - When using interest points or sampling: no guarantee to capture object-level parts ⇒ Dense sampling is often better.

Topics of This Lecture

- Recap: Specific Object Recognition with Local Features
- Part-Based Models for Object Categorization
  - Structure representations
  - Different connectivity structures
- Bag-of-Words Model
  - Use for image classification
- Implicit Shape Model
  - Generalized Hough Transform for object category detection
- Deformable Part-based Model
  - Discriminative part-based detection
Implicit Shape Model (ISM)

- **Basic ideas**
  - Learn an appearance codebook
  - Learn a star-topology structural model
    - Features are considered independent given obj. center

- **Algorithm: probabilistic Gen. Hough Transform**
  - Exact correspondences \( \rightarrow \) Prob. match to object part
  - NN matching \( \rightarrow \) Soft matching
  - Feature location on obj. \( \rightarrow \) Part location distribution
  - Uniform votes \( \rightarrow \) Probabilistic vote weighting
  - Quantized Hough array \( \rightarrow \) Continuous Hough space

Implicit Shape Model: Basic Idea

- Visual vocabulary is used to index votes for object position [a visual word = “part”].

**Training image**

Visual codeword with displacement vectors

Implicit Shape Model: Basic Idea

- Objects are detected as consistent configurations of the observed parts (visual words).

**Test image**


Implicit Shape Model - Representation

- Learn appearance codebook
  - Extract local features at interest points
  - Agglomerative clustering \( \rightarrow \) codebook
- Learn spatial distributions
  - Match codebook to training images
  - Record matching positions on object

Spatial occurrence distributions

Implicit Shape Model - Recognition

**Interest Points**

Matched Codebook Entries

Probabilistic Voting

Image Feature

Interpretation (Codebook match)

Object Position

3D Voting Space (continuous)

Probabilistic vote weighting

Example: Results on Cows

3rd hypothesis

Scale Invariant Voting

- Scale-invariant feature selection
  - Scale-invariant interest regions
  - Extract scale-invariant descriptors
  - Match to appearance codebook

- Generate scale votes
  - Scale as 3rd dimension in voting space
  - Search for maxima in 3D voting space

Detection Results

- Qualitative Performance
  - Recognizes different kinds of objects
  - Robust to clutter, occlusion, noise, low contrast

Detections Using Ground Plane Constraints

Battery of 5 ISM detectors for different car views

Extension: Rotation-Invariant Detection

- Polar instead of Cartesian voting scheme

- Benefits:
  - Recognize objects under image-plane rotations
  - Possibility to share parts between articulations.

- Caveats:
  - Rotation invariance should only be used when it’s really needed.
  - (Also increases false positive detections)

Sometimes, Rotation Invariance Is Needed...
**Implicit Shape Model - Segmentation**

- Local Features
- Matched Codebook Entries
- Probabilistic Voting
- Backprojected Hypotheses
- Backprojection of Maxima
- 3D Voting Space (continuous)

**Example Results: Motorbikes**

**Topics of This Lecture**

- Recap: Specific Object Recognition with Local Features
- Part-Based Models for Object Categorization
  - Structure representations
  - Different connectivity structures
- Bag-of-Words Model
  - Use for image classification
- Implicit Shape Model
  - Generalized Hough Transform for object category detection
- Deformable Part-based Model
  - Discriminative part-based detection

**You Can Try It At Home...**

- Linux source code & binaries available
  - Including datasets & several pre-trained detectors
  - [http://www.vision.rwth-aachen.de/software](http://www.vision.rwth-aachen.de/software)

**Starting Point: HOG Sliding-Window Detector**

- Array of weights for features in window of HOG pyramid
- Score is dot product of filter and vector

**Deformable Part-based Models**

- Mixture of deformable part models (pictorial structures)
- Each component has global template + deformable parts
- Fully trained from bounding boxes alone
2-Component Bicycle Model

- Root filters (coarse resolution)
- Part filters (finer resolution)
- Deformation models

Object Hypothesis

- Multi-scale model captures features at two resolutions

Score of an Hypothesis

\[
\text{score}(p_0, \ldots, p_n) = \sum_{i=0}^{m} F_i \cdot \phi(H, p_i) - \sum_{i=1}^{n} d_i \cdot (dx^2_i, dy^2_i)
\]

\[
\text{score}(z) = \beta \cdot \Psi(H, z)
\]

Score of Object Hypothesis

Score of filter: dot product of filter with HOG features beneath it

Score of object hypothesis is sum of filter scores minus deformation costs

Recognition Model

\[
f_w(x) = w \cdot \Phi(x)
\]

\[
f_w(x) = \max_z w \cdot \Phi(x, z)
\]

- \(z\): vector of part offsets
- \(\Phi(x, z)\): vector of HOG features (from root filter & appropriate part sub-windows) and part offsets

Results: Persons

- Results (after non-maximum suppression)
  - 1s to search all scales

Results: Bicycles

- Results adapted from Trevor Darrell
False Positives

- Bicycles

Results: Cats

High-scoring true positives

High-scoring false positives (not enough overlap)

Slide credit: Pedro Felzenszwalb

You Can Try It At Home...

- Deformable part-based models have been very successful at several recent evaluations.
  - Currently, state-of-the-art approach in object detection
  - Source code and models trained on PASCAL 2006, 2007, and 2008 data are available here:

References and Further Reading

- Details about the ISM approach can be found in
  - Details about the DPMs can be found in
  - Try the ISM Linux binaries
    - [http://www.vision.ee.ethz.ch/bleibe/code](http://www.vision.ee.ethz.ch/bleibe/code)
  - Try the Deformable Part-based Models
    - [http://www.cs.uchicago.edu/~pff/latent](http://www.cs.uchicago.edu/~pff/latent)