Object Recognition with Deformable Models

Pedro F. Felzenszwalb
Department of Computer Science
University of Chicago

Joint work with: Dan Huttenlocher, Joshua Schwartz,
David McAllester, Deva Ramanan.
Example Problems

Detecting rigid objects

Detecting non-rigid objects

PASCAL challenge

Medical image analysis

Segmenting cells
Deformable Models

• Significant challenge:
  - Handling variation in appearance within object classes
  - Non-rigid objects, generic categories, etc.

• Deformable models approach:
  - Consider each object as a deformed version of a template
  - Compact representation
  - Leads to interesting modeling and algorithmic problems
Overview

• **Part I: Pictorial Structures**
  - Deformable part models
  - Highly efficient matching algorithms

• **Part II: Deformable Shapes**
  - Triangulated polygons
  - Hierarchical models

• **Part III: The PASCAL Challenge**
  - Recognizing 20 object categories in realistic scenes
  - Discriminatively trained, multiscale, deformable part models
Part I: Pictorial Structures

- Introduced by Fischler and Elschlager in 1973
- Part-based models:
  - Each part represents local visual properties
  - “Springs” capture spatial relationships

Matching model to image involves joint optimization of part locations “stretch and fit”
Local Evidence + Global Decision

- Parts have a match quality at each image location
- Local evidence is noisy
  - Parts are detected in the context of the whole model
Matching Problem

- Model is represented by a graph $G = (V, E)$
  - $V = \{v_1, \ldots, v_n\}$ are the parts
  - $(v_i, v_j) \in E$ indicates a connection between parts
- $m_i(l_i)$ is a cost for placing part $i$ at location $l_i$
- $d_{ij}(l_i, l_j)$ is a deformation cost
- Optimal configuration for the object is $L = (l_1, \ldots, l_n)$ minimizing

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

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

- Assume \( n \) parts, \( k \) possible locations for each part
  - There are \( k^n \) configurations \( L \)
- If graph is a tree we can use dynamic programming
  - \( O(nk^2) \) algorithm
- If \( d_{ij}(l_i,l_j) = g(l_i-l_j) \) we can use min-convolutions
  - \( O(nk) \) algorithm
  - As fast as matching each part separately!
Human Pose Estimation
Human Tracking

Ramanan, Forsyth, Zisserman, *Tracking People by Learning their Appearance*  
Part III: PASCAL Challenge

- ~10,000 images, with ~25,000 target objects
  - Objects from 20 categories (person, car, bicycle, cow, table...)
  - Objects are annotated with labeled bounding boxes
Model Overview

Model has a root filter plus deformable parts
Histogram of Gradient (HOG) Features

- Image is partitioned into 8x8 pixel blocks
- In each block we compute a histogram of gradient orientations
  - Invariant to changes in lighting, small deformations, etc.
- We compute features at different resolutions (pyramid)
Filters

- Filters are rectangular templates defining weights for features
- Score is dot product of filter and subwindow of HOG pyramid

![Diagram showing the process from an image pyramid to a HOG pyramid with a filter applied](image)

Score of $H$ at this location is $H \cdot W$
Object Hypothesis

Multiscale model captures features at two-resolutions

Score is sum of filter scores plus deformation scores
Training

- Training data consists of images with labeled bounding boxes
- Need to learn the model structure, filters and deformation costs
Learned Models

Bottle

Car

Sofa

Bicycle
Example Results
More Results
Overall Results

• 9 systems competed in the 2007 challenge
• Out of 20 classes we get:
  - First place in 10 classes
  - Second place in 6 classes
• Some statistics:
  - It takes ~2 seconds to evaluate a model in one image
  - It takes ~3 hours to train a model
  - MUCH faster than most systems
Summary

• Deformable models provide an elegant framework for object detection and recognition
  - Efficient algorithms for matching models to images
  - Applications: pose estimation, medical image analysis, object recognition, etc.

• We can learn models from partially labeled data
  - Generalized standard ideas from machine learning
  - Leads to state-of-the-art results in PASCAL challenge

• Future work: hierarchical models, grammars, 3D objects