Object Recognition and Localization

Andrew Zisserman
Visual Geometry Group
University of Oxford
http://www.robots.ox.ac.uk/~vgg

Includes slides from: Anna Bosch, Ondra Chum, Mark Everingham, Rob Fergus, Pawan Kumar, Bastian Leibe, Pietro Perona, Bernt Schiele, Josef Sivic and Manik Varma
What we would like to be able to do …

- Visual recognition and scene understanding
- *What* is in the image and *where*

- scene type: outdoor, city …
- object classes
- material properties
- actions
Various tasks

Image classification:
“the image contains an airplane”

Object detection/localization:
“the object is here, in a bounding box”

Object pixel-level segmentation:
“the object ‘owns’ these pixels”
Why is the recognition problem hard?

- Scale and shape of the imaged object varies with viewpoint
- Occlusion (self- or by a foreground object)
- Lighting changes
- Background “clutter”
Some object classes (Caltech datasets)

Difficulties:

- Size/shape variation
- Partial occlusion
- Lighting
- Background clutter
- Intra-class variation
Class of model: Pictorial Structure

- Intuitive model of an object
- Model has two components
  1. parts (2D image fragments)
  2. structure (configuration of parts)
- Dates back to Fischler & Elschlager 1973

Is this complexity of representation necessary?
Deformations
Learning

Unclear how to model categories, so learn rather than manually specify
Levels of supervision in learning

none (weak supervision)

ROI

“parts”

pixel level (regions)
Specific object recognition history:
Geometric methods

David Lowe [1985]

Rothwell et al. [1992]
Specific object recognition history: Appearance-based methods

- Murase & Nayer 1995
- Schmid & Mohr 1996
- Lowe 1999
- ...

Original use of SIFT descriptor
History: early object categorization

- LeCun et al. 1998
- Amit and Geman, 1999
- Belongie and Malik, 2002
- DeCoste and Scholkopf, 2002
- Simard et al. 2003

- Rowley & Kanade, 1998
- Schneiderman & Kanade 2000
- Viola and Jones, 2000
- Heisele et al., 2001

- Poggio et al. 1993
- Schneiderman & Kanade, 2000
- Argawal and Roth, 2002

....
Application I: improving photography

[Face priority AE] When a bright part of the face is too bright
Application II: Improving online search

Query: STREET

Application III: Organizing photo collections
Outline

1. Bag of visual words model for categorization
   • SVM classifier

2. Adding spatial information for localization

3. Databases and challenges

4. Spatial layout

5. Class based segmentation
   • Pixel level localization

6. Conclusions and the future