Introduction to Object Recognition

Achim J. Lilienthal

AASS, Dept. Technology, Örebro University

Docent Lecture, Computer Vision 2008-04-08
The Problem from a Computer's Perspective
Image = An Array of Numbers
Object Recognition In the Sea of Numbers

contains a cup

cup handle
1. Defining the Subject

2. Object Recognition – Sub-Tasks

3. Object Recognition – Challenges

4. Object Recognition – Standard Approaches

5. The "Did You Mean?" Approach

6. Summary
Defining the Subject
1 Introduction – Objects

- Philosophy
  - a thing, an entity or a being

- Non-abstract Objects
  - things located somehow in space and time

1 Introduction – Objects

- Philosophy
  - a thing, an entity or a being

- Non-abstract Objects
  - things located somehow in space and time
  - something **perceptible** by one or more of the senses, especially by **vision** or touch; a material thing

1 Introduction – Describing Objects

Objects as Properties and Relations

example: cup

properties

• colour, texture
• size
• composition

relations

• standing on a saucer
• contains a liquid
• smaller than a bucket
Object Recognition
Sub-Tasks
2 Object Recognition – Tasks

- Verification
  - Is this a lamp?

Fei-Fei et al., CVPR 2007, Short Course
2  Object Recognition – Tasks

- Detection
  incl. Localisation
  - Are there people and where?
2 Object Recognition – Tasks

- Identification
  - Is that Potala Palace?
2 Object Recognition – Tasks

- Object Categorization

- Image of a street scene with labeled objects:
  - palace
  - mountain
  - tree
  - building
  - street lamp
  - garbage bin
- People visible in the scene.
2 Object Recognition – Tasks

- Scene/Context Categorization

- outdoor
- city
- ...
Object Recognition
Challenges
3 Object Recognition in Computer Vision

- Object Recognition in Computer Vision
- images as an array of numbers
3 Challenges

The Core Problem

- wide real-world image variation
3 Challenges

- View Point Variation

Why is Real-World Visual Object Recognition Hard?
3 Challenges

- Illumination
3 Challenges

- Occlusion

Magritte, 1957
3 Challenges

- Scale
3 Challenges

- Deformation
Object Recognition
Standard Approaches
Object Recognition as a Classification Problem

- image is partitioned into a set of overlapping windows
- decision at each window whether it contains the object
- classification function is learnt from examples

**where are the screens?**

look at a set of image patches
Example: Face Detection
- how many faces?
4 Object Recognition – Standard Approaches

- Object Recognition as a Classification Problem
  - searching over a range of scales
Object Recognition as a Classification Problem

- searching over a range of scales

- state of the art face detector
  - suitable parameter value

*Face Detection - Efficient and Rank Deficient.*
4 Object Recognition – Standard Approaches

- Object Recognition as a Classification Problem
  - searching over a range of scales
  - state of the art face detector
    - unsuitable parameter value

*Face Detection - Efficient and Rank Deficient.*
Object Recognition as a Classification Problem

- searching over a range of
  - scales
  - orientations
4 Object Recognition – Standard Approaches

- Object Recognition as a Classification Problem
  - searching over a range of
    - scales
    - orientations

*Face Detection - Efficient and Rank Deficient.*
4 Representation

Representation of the Small Image Patches?
4 Representation

Local Descriptors, Features

- find salient points
  - interest points:
    "good" to recognize despite translation, rotation, scale, ...
    (invariance)
- describe interest points
  - invariance
  - discriminative
    (not confused with other features)

\[ \rightarrow [x_1, x_2, x_3, x_4, x_5, \ldots, x_N] \]
4 Representation

- Object Recognition with Local Descriptors

David Lowe, Computer Vision Course, 2007
4 Representation

- Representation of Feature Sets
4 Representation

- Representation of Sets
  - bag of words models
4 Representation

- Representation of Feature Sets – Structural Models
  - part-based models
The "Did You Mean?" Approach
5 The "Did You Mean?" Approach

- Google's "Did you mean?" Tool
  - a large dataset allows for simple algorithms
    - no complex parsing / image processing
    - memorizing of billions of queries pairs
    - suggesting the one closest to the user's query

![Google Search](image)
A Large Dataset for Object and Scene Recognition

- a fundamentally different approach
  1. Internet is huge, contains billions of images
  2. collect a large database of labelled images
  3. simple comparison of the query image with the images in the database
  4. assign the label of the visually closest images found

A Large Dataset for Object and Scene Recognition

a fundamentally different approach

1. Internet is huge, contains billions of images
2. collect a large database of labelled images
3. simple comparison of the query image with the images in the database
4. assign the label of the visually closest images found

Discuss in groups of two (or three) [3 min]

- what are the main difficulties of this approach?
5 A Large Dataset for Object Recognition

A Large Dataset for Object and Scene Recognition

difficulty 1 – how to collect a large labelled dataset?
difficulty 2 – how many labels? where to get them from?
difficulty 3 – how large should the images be?
difficulty 4 – how many images?
difficulty 5 – how to compare images?
difficulty 6 – aren't there many falsely labelled images?

Method

part 1 → part 2 → summary → performance
5 A Large Dataset for Object Recognition

A Large Dataset for Object and Scene Recognition

difficulty 1 – collecting a labelled dataset

- ask the community to label images
  - LabelMe
    - http://www.csail.mit.edu/~brussell/research/LabelMe/intro.html
  - currently around 40000 labelled images (March 2008)
difficulty 1 – collecting a labelled dataset

use image search engines

- Altavista, Ask, Flickr, Cydral, Google, Picsearch, Webshots
- collect images for all non-abstract nouns

A person icon is used as a rough
364 x 769 - 5k - gdf
www.veerstan.org

Donald R. Perone, Esq.
375 x 450 - 95k - jpg
www.jamesdeter.com

... "missing person" 寄人疏
361 x 600 - 73k - jpg
www.chinesejediblog.com

... is at Carl B. Personer in Costume Houston 460 x 640 - 146k - jpg
www.lawinattorneygeneral.com

... person close to star says
300 x 305 - 52k - jpg
www.cbc.ca

A small lonely person in a big
world...
512 x 494 - 103k - jpg
www.civil liberties.org

... person Bike from Hanover ...
400 x 340 - 48k - jpg
www.gadgetpage.com

I don't like it when people say ...
1500 x 1500 - 3900k - jpg
bustedbybuildidlers.blogspot.com

Person at the Window Art Print by
322 x 425 - 28k - jpg
www.allposters.com
difficulty 2 – how many labels? where to get them from?

- 75'846 English non-abstract nouns from Wordnet
- Wordnet provides class names and the semantic relations between the classes

WordNet: a lexical database for the English language

Method

- substantial image download
  - 8 months (max. 3000 images per noun)
  - approx. 80 million unique images
  - loosely labelled with 75,000 English non-abstract nouns from Wordnet
difficulty 3 – how large should the images be?

- only $16 \times 16$ pixels required for robust face detection
- experimental evidence suggests that $32 \times 32$ pixel is sufficient for scene recognition and object detection

.Masking in Visual Recognition: Effects of Two-Dimensional Noise,
difficulty 4 – how many images?

- effectively there is an infinite number of images
  - $32 \times 32$ pixels means $32 \times 32 \times 3 = 3072$ dimensions
  - each dimension having 8 bits there are $10^{7400}$ possible images
  - human in 100 years only gets to see $10^{11}$ frames (30 frames/s)

- but the visual world is very regular
  - (real world pictures occupy only a relatively small portion of the space of possible images)

- images must cover the manifold of natural images densely

- as many images as possible
  - (given the memory, processing time constraints)
5 A Large Dataset for Object Recognition

Method

- substantial image download $\rightarrow$ tiny image database
  - 8 months (max. 3000 images per noun)
  - approx. 80 million unique images
  - loosely labelled with 75'000 English non-abstract nouns from Wordnet
  - $32+ \times 32+$ colour pixels (32 pixels on the smallest axis)
  - approx. 760 Gb (parts available at http://people.csail.mit.edu/torralba/tinyimages/)
A Large Dataset for Object and Scene Recognition

- difficulty 5 – how to compare images?
- simplest way
  - sum of squared differences (SSD) between the 3072 dim. vectors (each image is normalized with zero mean and unit norm)
A Large Dataset for Object and Scene Recognition

- difficulty 5 – how to compare images?
- simplest way
  - sum of squared differences (SSD) between the 3072 dim. vectors (each image is normalized with zero mean and unit norm)
A Large Dataset for Object Recognition

- difficulty 5 – how to compare images?
- more complex way with better performance
  - allow limited translation, scaling and mirroring to minimize SSD

Target  Neighbor  Warping  Pixel shifting
A Large Dataset for Object and Scene Recognition

difficulty 6 – aren't there many falsely labelled images?

- yes, there is a lot of labeling noise (*images are loosely labelled*)
  around 40% of the first 30 images are correct on average

- Google and Altavista performed the best

- methods works despite the high labelling noise
  (using a voting scheme)
Method

- substantial image download → tiny image database
  - 8 months (max. 3000 images per noun)
  - approx. 80 million unique images
    - loosely labelled with 75'000 English non-abstract nouns
  - \( 32^+ \times 32^+ \) colour pixels (32 pixels on the smallest axis)

- simple image comparison
  - evaluate correlation directly (plus some tricks!)
  - determine next neighbours
  - use a voting scheme (K-nearest neighbour approach)
5 A Large Dataset for Object Recognition

Object / Scene Recognition Performance
5 A Large Dataset for Object Recognition

Object / Scene Recognition Performance
Summary
6 Summary

1. Defining the Subject
   - object recognition
   - object recognition in computer vision
6 Summary

1. Defining the Subject

2. Object Recognition – Sub-Tasks
6 Summary

1. Defining the Subject
2. Object Recognition – Sub-Tasks
3. Object Recognition – Challenges
6 Summary

1. Defining the Subject
2. Object Recognition – Sub-Tasks
3. Object Recognition – Challenges
4. Object Recognition – Standard Approaches
6 Summary

1. Defining the Subject
2. Object Recognition – Sub-Tasks
3. Object Recognition – Challenges
4. Object Recognition – Standard Approaches
5. The "Did You Mean?" Approach
Future Work
7 Future Work

- Representation
  - SIFT, SURF, ...

- Learning
  - Neural Networks, SVM, AdaBoost, ...

- Collection of Datasets

- Top-Down Approaches
  - analysis-by-synthesis

- Context-Sensitive Object Recognition
7 Future Work

- Role of Context?
  - typical problems with local detectors

Torralba et al.,
CVPR 2007, Short Course
7 Future Work

- Role of Context?
  - we as humans use context
7 Future Work

- Role of Context?
  - we as humans use context
Thank you for your attention!
Waffles!