CMPSCI 370: Intro. to Computer Vision

Introduction to recognition

University of Massachusetts, Amherst
March 29, 2014

Instructor: Subhransu Maji

Announcements

- Midterm has been graded
  - Average score: **54.2** (out of 80)
  - Come by my office hours
    - if you have any questions
    - or did not collect the midterm in class
    - or to chat about the latest AI technology (AlphaGo, Holoportation, ….)
- Homework 3 grades will be available shortly
- **No class** this Thursday (3/31) due to instructor's travel
- **No honors section** today

Common mistakes …

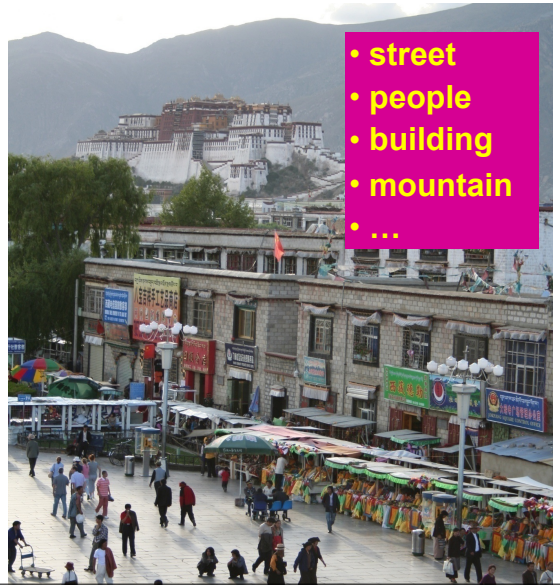
- What is a Bayer filter for?
  - "for image smoothing"
  - "for color sensing in digital cameras"
- A technique to enhance the contrast of an image:
  - "sharpen the image" — sharpening is not the same as contrast enhancement
  - "gamma/log-normalization", "brightness stretching", "histogram equalization"
- Factor that lead to edges
  - "gx, gy is high"
  - "occlusion, shadows, change in surface orientation, texture,…"

Object Recognition: Overview and History

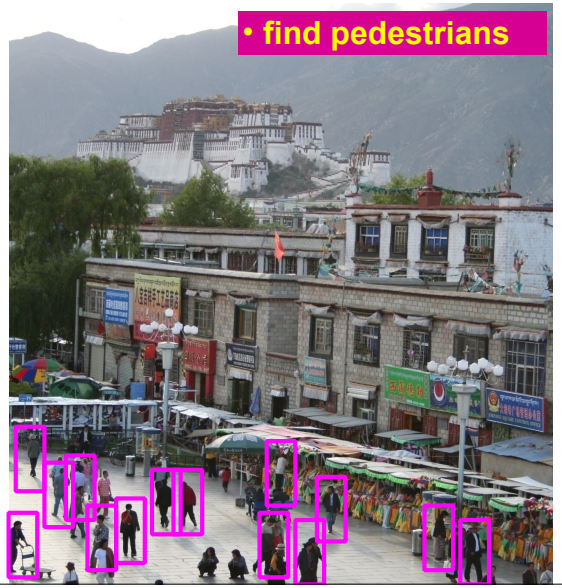
Slides adapted from Svetlanan Lazebnik, Alex Berg, Fei-Fei Li, Rob Fergus, Antonio Torralba, and Jean Ponce

Scene categorization

Image annotation/tagging



Object detection



Activity recognition

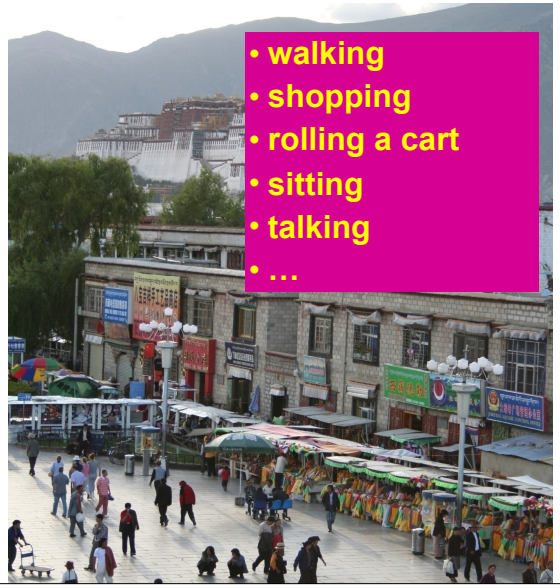
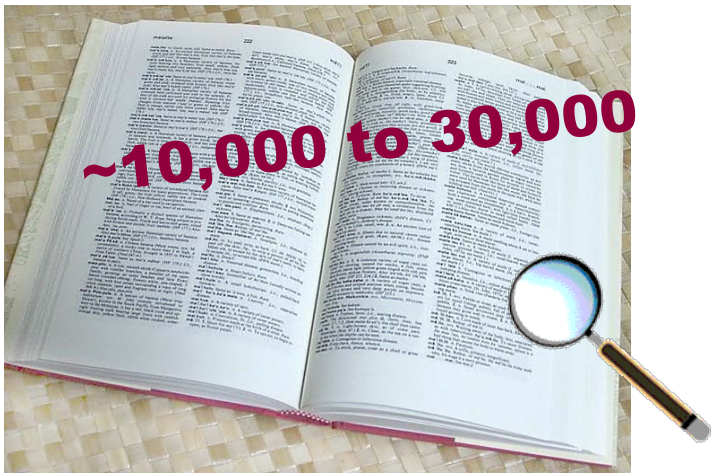


Image parsing

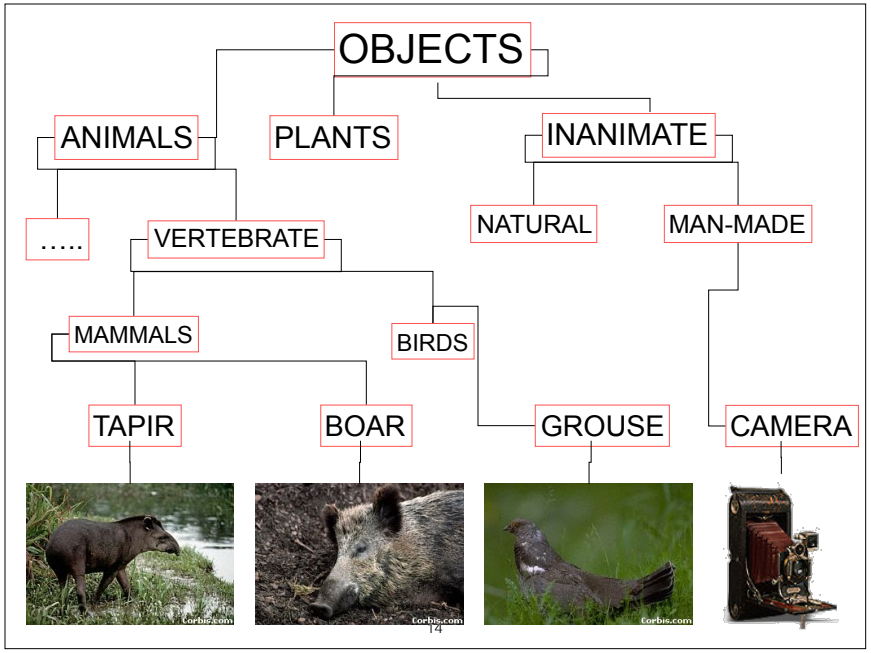
Image understanding?

How many visual object categories are there?

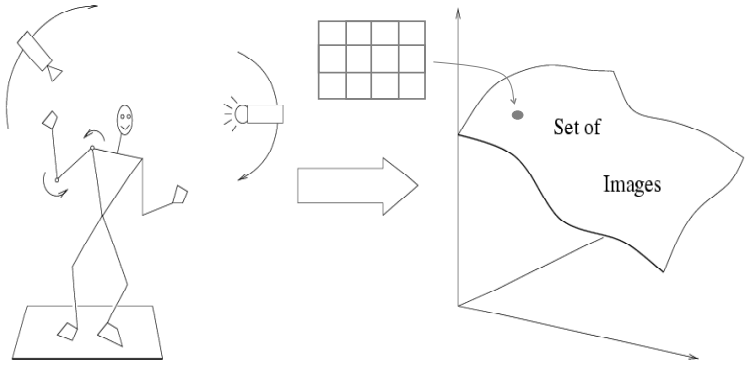


http://wexler.free.fr/library/files/biederman%20(1987)%20recognition-by-components.%20a%20theory%20of%20human%20image%20understanding.pdf

Biederman 1987



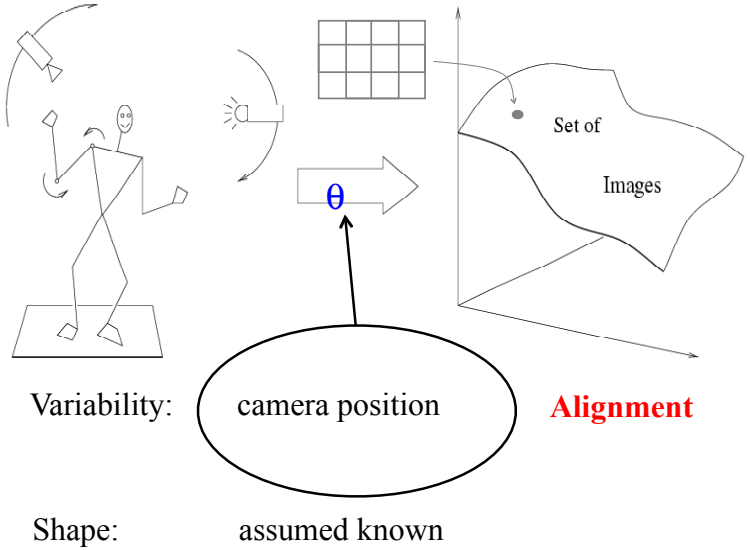
Recognition is all about modeling variability



Variability: Camera position
Illumination
Within-class variation
Background, occlusion

History of ideas in recognition

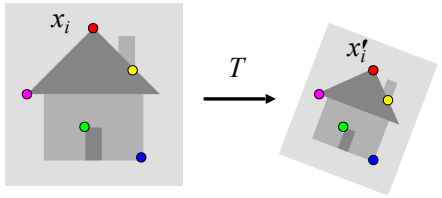
1960s – early 1990s: the geometric era



Roberts (1965); Lowe (1987); Faugeras & Hebert (1986); Grimson & Lozano-Perez (1986); Huttenlocher & Ullman (1987)

Recall: Alignment

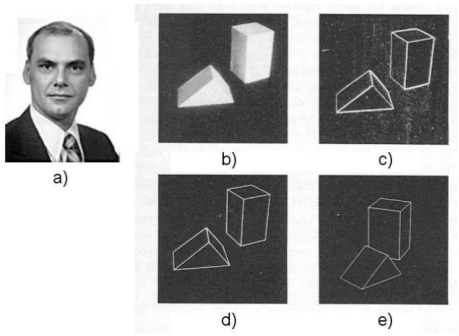
Alignment: fitting a model to a transformation between pairs of features (*matches*) in two images



Find transformation T that minimizes

$$\sum_i \text{residual}(T(x_i), x_i')$$

Recognition as an alignment problem: Block world

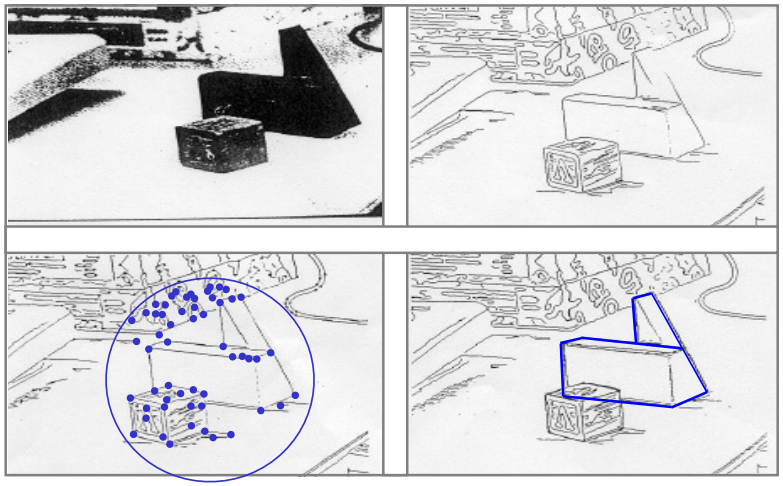


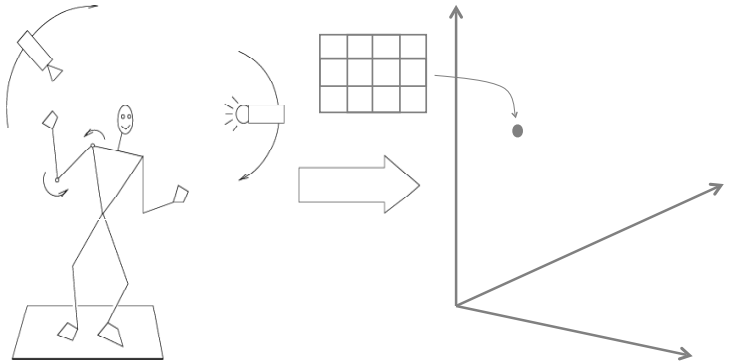
L. G. Roberts, ***Machine Perception of Three Dimensional Solids***, Ph.D. thesis, MIT Department of Electrical Engineering, 1963.

Fig. 1. A system for recognizing 3-d polyhedral scenes. a) L.G. Roberts. b)A blocks world scene. c)Detected edges using a 2x2 gradient operator. d) A 3-d polyhedral description of the scene, formed automatically from the single image. e) The 3-d scene displayed with a viewpoint different from the original image to demonstrate its accuracy and completeness. (b) - e) are taken from [64] with permission MIT Press.)

J. Mundy, **Object Recognition in the Geometric Era: a Retrospective**, 2006

Alignment: Huttenlocher & Ullman (1987)



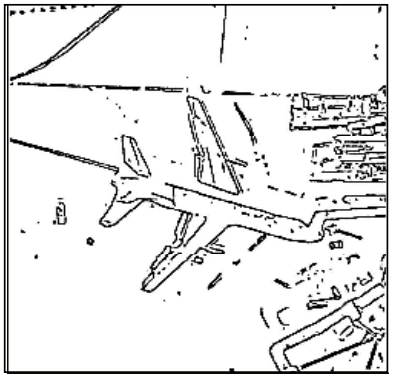


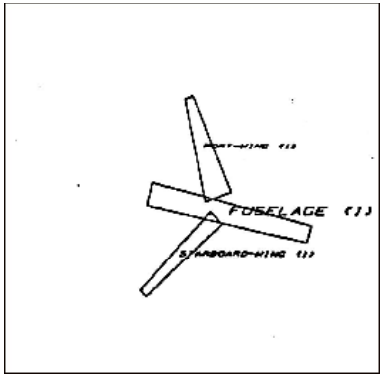
~~Variability~~

Invariance to: Camera position
Illumination
Etc.

Duda & Hart (1972); Weiss (1987); Mundy et al. (1992-94); Rothwell et al. (1992); Burns et al. (1993)

From object instances to object categories

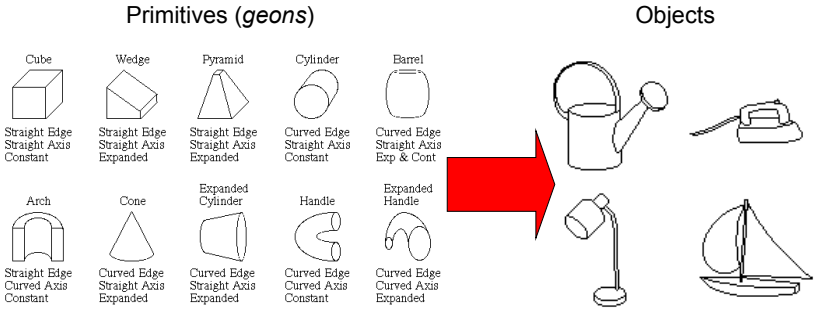




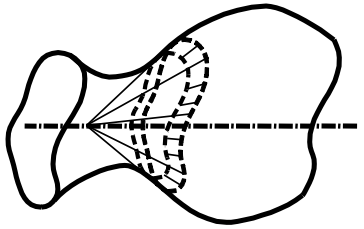
ACRONYM (Brooks and Binford, 1981)
Binford (1971), Nevatia & Binford (1972), Marr & Nishihara (1978)

Recognition by components

Biederman (1987)

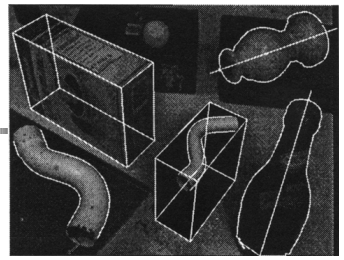


http://en.wikipedia.org/wiki/Recognition_by_Components_Theory

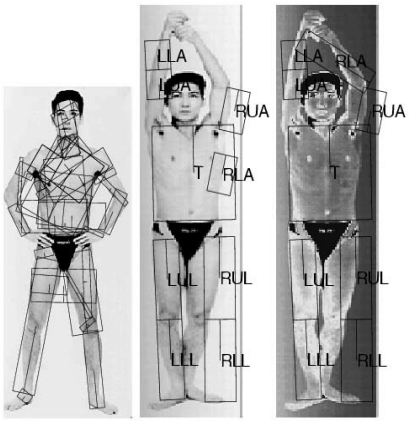


Generalized cylinders
Ponce et al. (1989)

General shape primitives?



Zisserman et al. (1995)

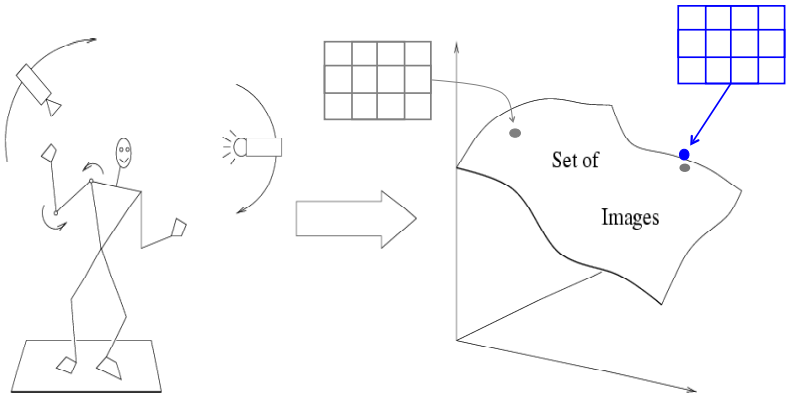


Forsyth (2000)

History of ideas in recognition

1960s – early 1990s: the geometric era

1990s: appearance-based models

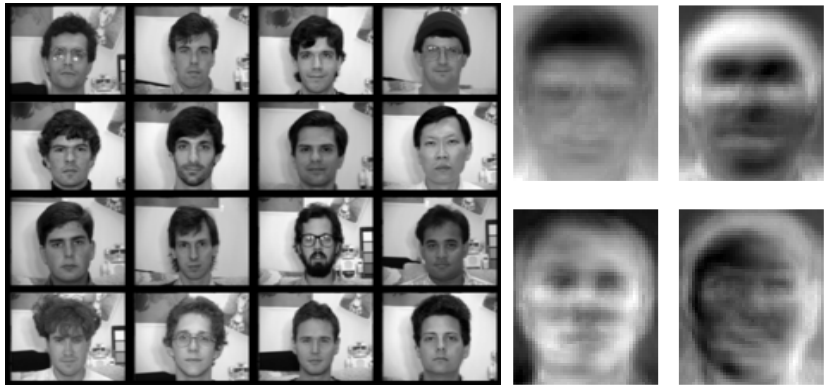


Empirical models of image variability

Appearance-based techniques

Turk & Pentland (1991); Murase & Nayar (1995); etc.

Eigenfaces (Turk & Pentland, 1991)

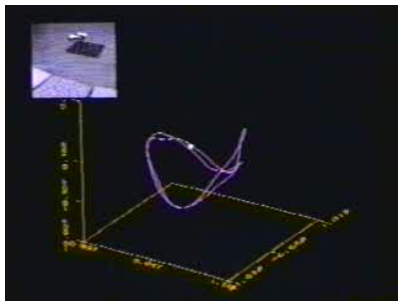


Experimental	Correct/Unknown Recognition Percentage		
Condition	Lighting	Orientation	Scale
Forced classification	96/0	85/0	64/0
Forced 100% accuracy	100/19	100/39	100/60
Forced 20% unknown rate	100/20	94/20	74/20

Color Histograms

Swain and Ballard, **Color Indexing**, IJCV 1991.

Appearance manifolds



H. Murase and S. Nayar, Visual learning and recognition of 3-d objects from appearance, IJCV 1995

Limitations of global appearance models

Requires global registration of patterns

Not robust to clutter, occlusion, geometric transformations

History of ideas in recognition

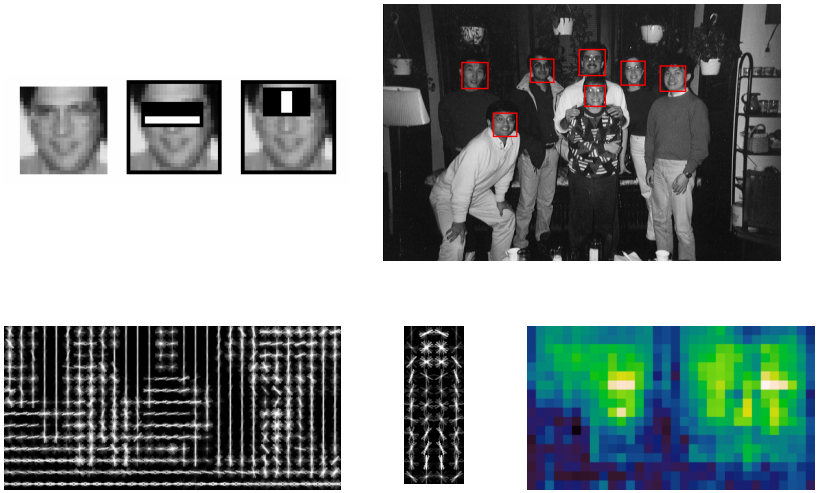
1960s – early 1990s: the geometric era

1990s: appearance-based models

1990s – present: sliding window approaches

Sliding window approaches

Sliding window approaches



History of ideas in recognition

1960s – early 1990s: the geometric era

1990s: appearance-based models

1990s – present: sliding window approaches

Late 1990s: local features

Local features for object instance recognition

D. Lowe (1999, 2004)

Large-scale image search

Combining local features, indexing, and spatial constraints

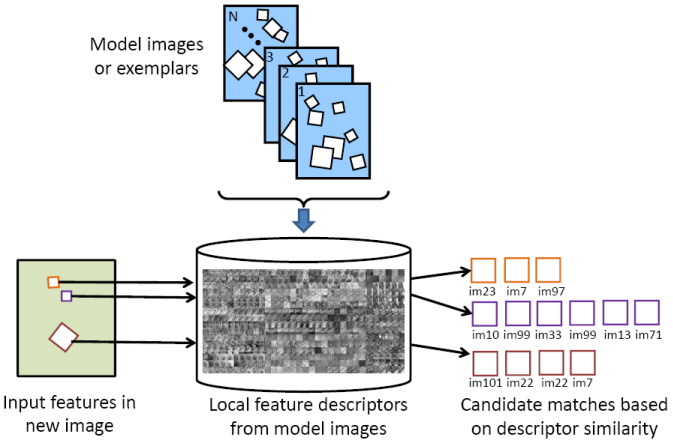


Image credit: K. Grauman and B. Leibe

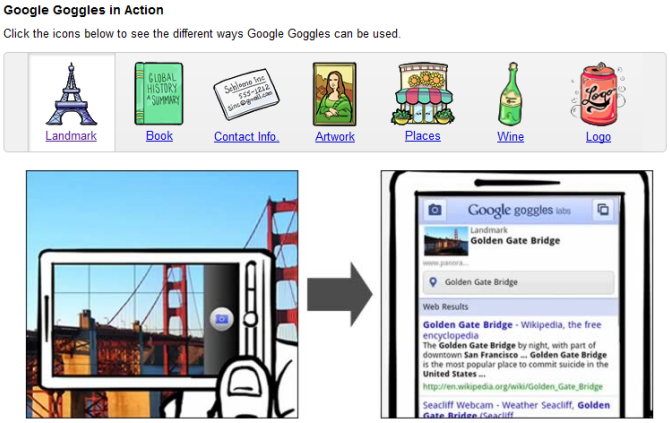
Large-scale image search

Combining local features, indexing, and spatial constraints

Philbin et al. ‘07

Large-scale image search

Combining local features, indexing, and spatial constraints



History of ideas in recognition

1960s – early 1990s: the geometric era

1990s: appearance-based models

1990s – present: sliding window approaches

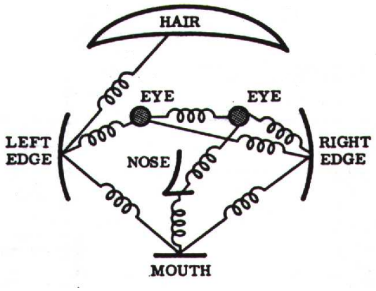
Late 1990s: local features

Early 2000s: parts-and-shape models

Parts-and-shape models

Model:

- Object as a set of parts
- Relative locations between parts
- Appearance of part



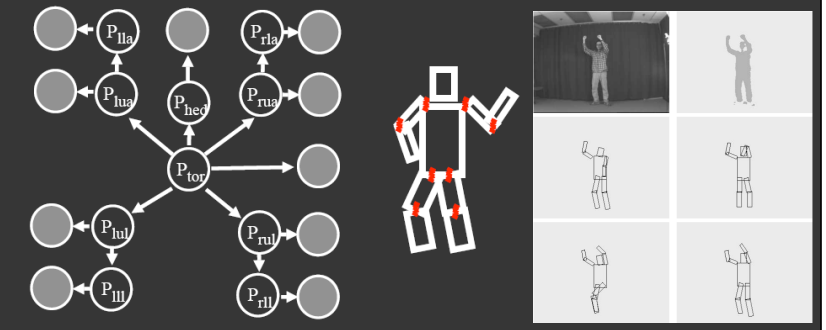
Fischler & Elschlager 73

Constellation models

Weber, Welling & Perona (2000), Fergus, Perona & Zisserman (2003)

Pictorial structure model

Fischler and Elschlager(73), Felzenszwalb and Huttenlocher(00)



$$\Pr(P_{tor}, P_{arm}, \ldots | \mathrm{Im}) \propto \prod_{i,j} \Pr(P_i | P_j) \prod_i \Pr(\mathrm{Im}(P_i))$$

part geometry

part appearance

History of ideas in recognition

1960s – early 1990s: the geometric era

1990s: appearance-based models

1990s – present: sliding window approaches

Late 1990s: local features

Early 2000s: parts-and-shape models

Mid/Late-2000s: bags of features, fully learned models

Bag-of-features models

Object → **Bag of 'words'**

Objects as texture

All of these are treated as being the same

No distinction between foreground and background: scene recognition?

Learning algorithms to the rescue.

Learned part-based models

Poselet detectors: Bourdev, Maji and Malik

Deformable part-based models, Girshick, Felzenszwalb, Ramanan, McAllester

History of ideas in recognition

1960s – early 1990s: the geometric era

1990s: appearance-based models

1990s – present: sliding window approaches

Late 1990s: local features

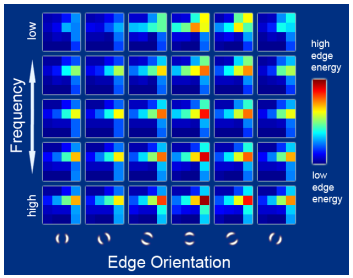
Early 2000s: parts-and-shape models

Mid-2000s: bags of features

Present trends: "big data", context, attributes, combining geometry and recognition, advanced scene understanding tasks, deep learning

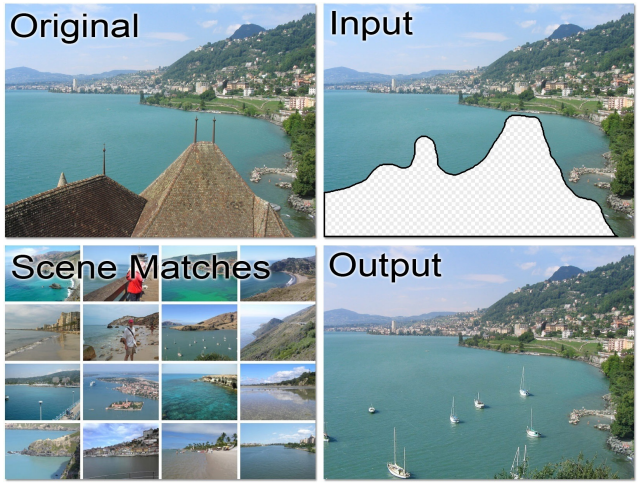
Global appearance models revisited

The "gist" of a scene: Oliva & Torralba (2001)



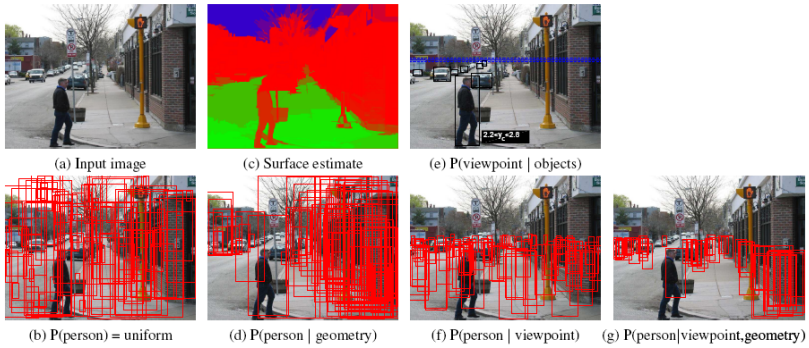
http://people.csail.mit.edu/torralba/code/spatialenvelope/

New applications in graphics



J. Hays and A. Efros, Scene Completion using Millions of Photographs, SIGGRAPH 2007

Geometric context

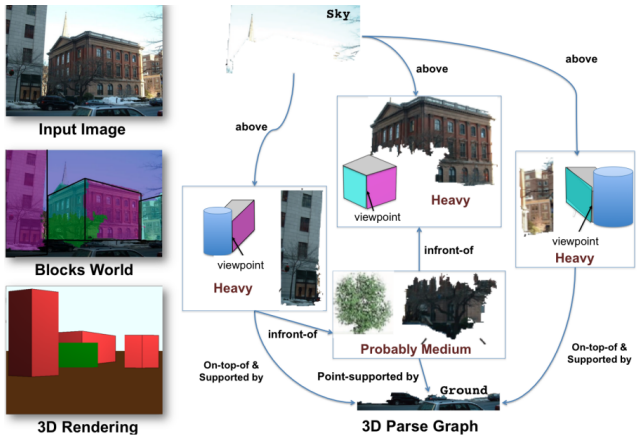


D. Hoiem, A. Efros, and M. Herbert, Putting Objects in Perspective, CVPR 2006

Geometry and recognition

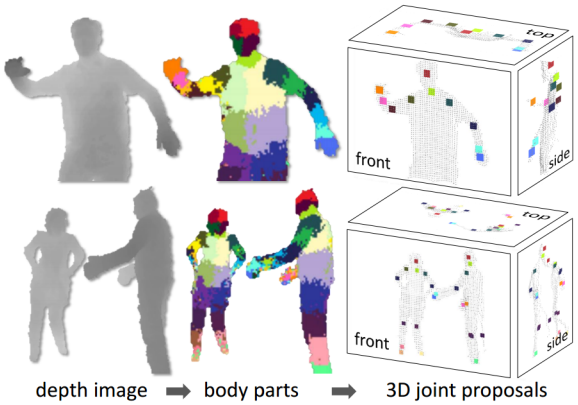
V. Hedau, D. Hoiem, and D. Forsyth, Recovering the Spatial Layout of Cluttered Rooms, ICCV 2009.

Geometry and recognition



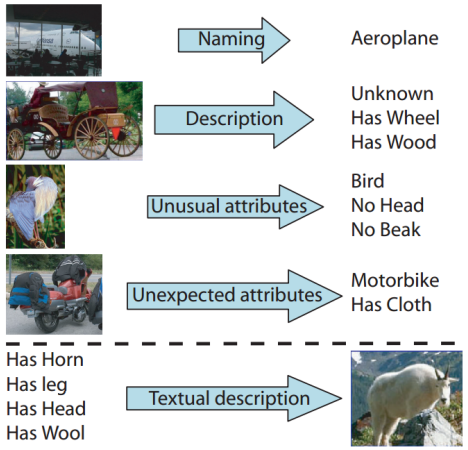
A. Gupta, A. Efros and M. Hebert, Blocks World Revisited: Image Understanding Using Qualitative Geometry and Mechanics, ECCV 2010

Recognition from RGBD Images



J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, **Real-Time Human Pose Recognition in Parts from a Single Depth Image**, CVPR 2011

Attributes for recognition

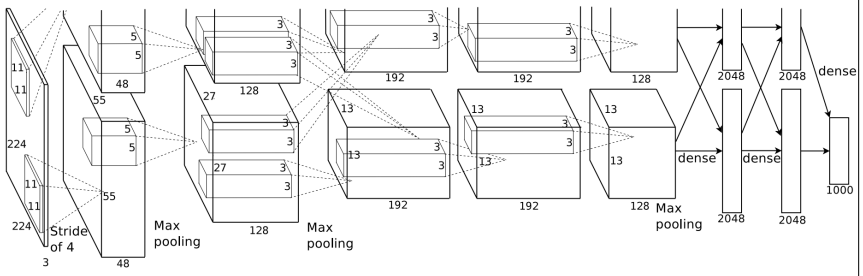


A. Farhadi, I. Endres, D. Hoiem, and D Forsyth, **Describing Objects by their Attributes**, CVPR 2009

Deep learning

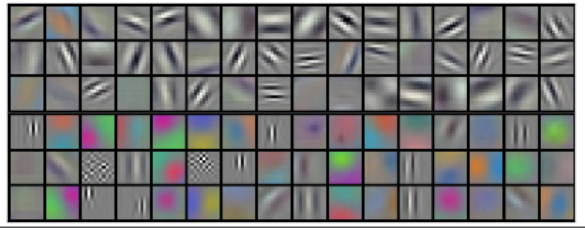
NY Times article

Recent deep learning breakthroughs…



ImageNet Classification with Deep Convolutional Neural Networks Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton NIPS 2014

96 filters learned in layer 1



Further thoughts and readings

- Chapter 14, Szeliski's book
- Think of the applications of computer vision around you