CMPSCI 670: Computer Vision

Introduction to recognition

University of Massachusetts, Amherst

October 27, 2014

Instructor: Subhransu Maji

Today

- **Administrivia:**
  - Project abstracts due today
    - email the pdf to me
  - Office hours ~~wednesday~~ today 3:45 - 4:45 pm (after class)
- **Outline** for today/tomorrow's class
  - What are the recognition problems in computer vision?
  - A historic perspective of methods

Object Recognition: Overview and History

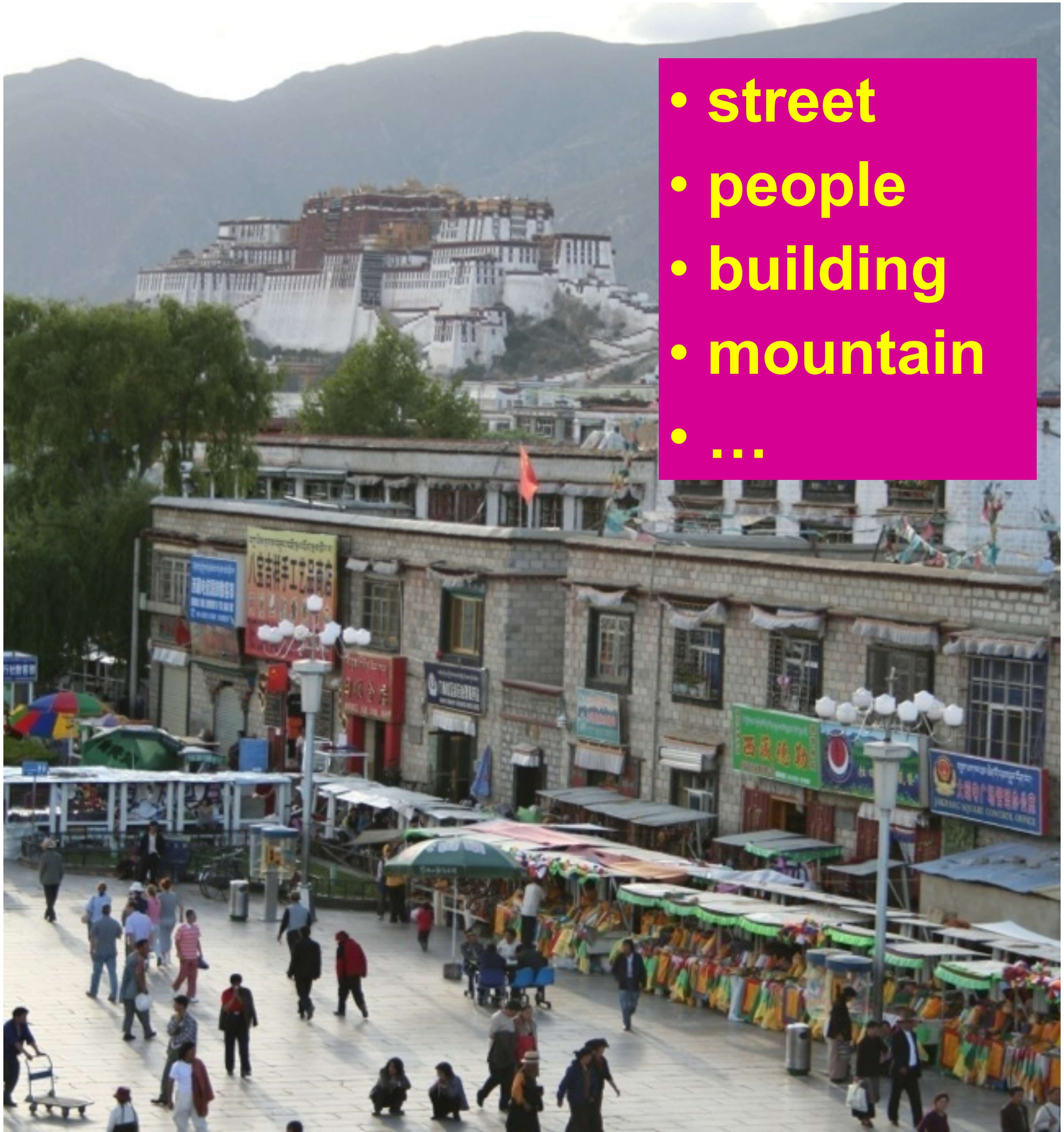
Slides adapted from Svetlanan Lazebnik, Alex Berg, Fei-Fei Li, Rob Fergus, Antonio Torralba, and Jean Ponce


JOKHANG SQUARE CONTROL OFFICE

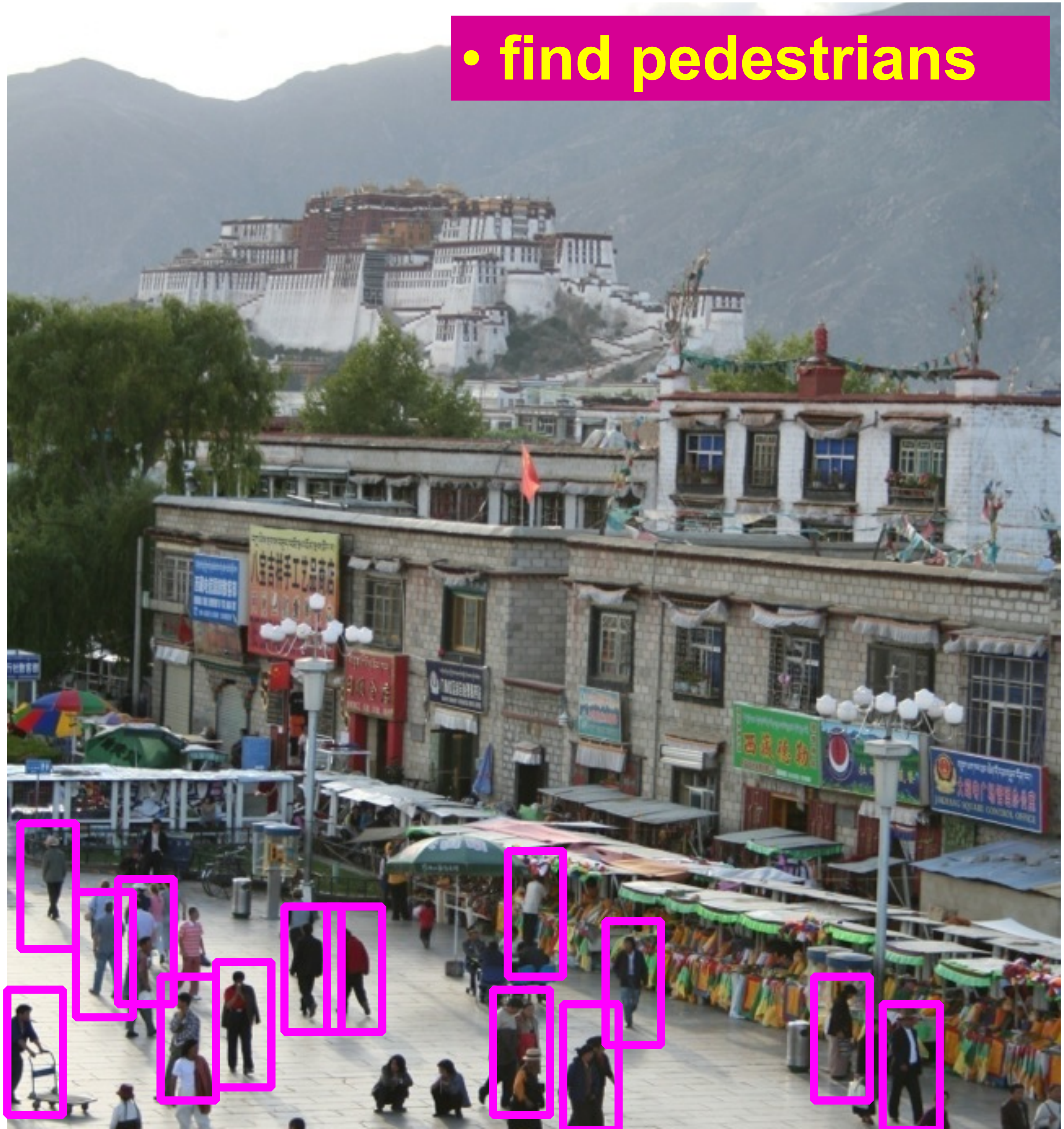
Scene categorization

- **outdoor/indoor**
- **city/forest/factory/etc.**

Image annotation/tagging



Object detection



Activity recognition

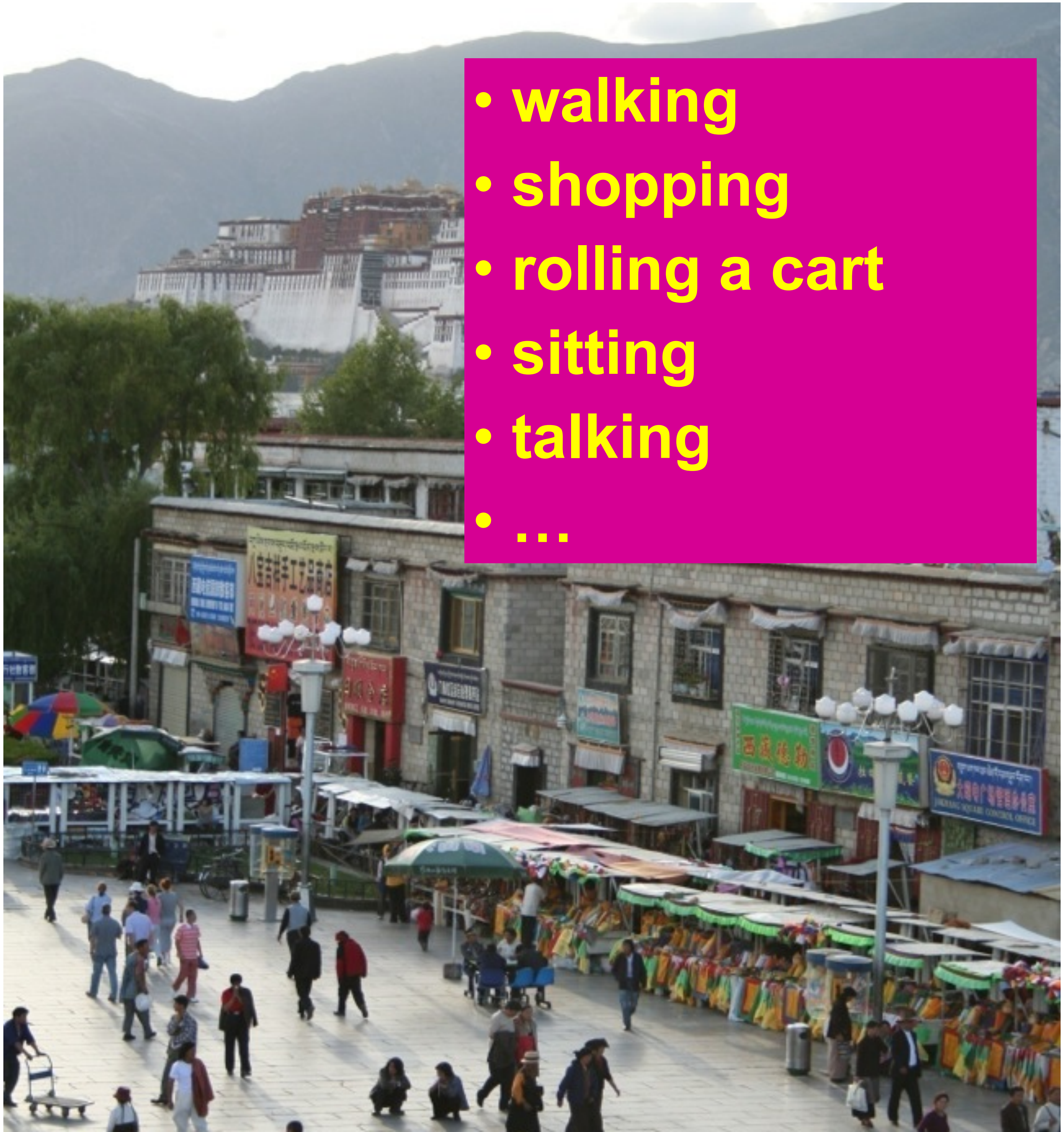
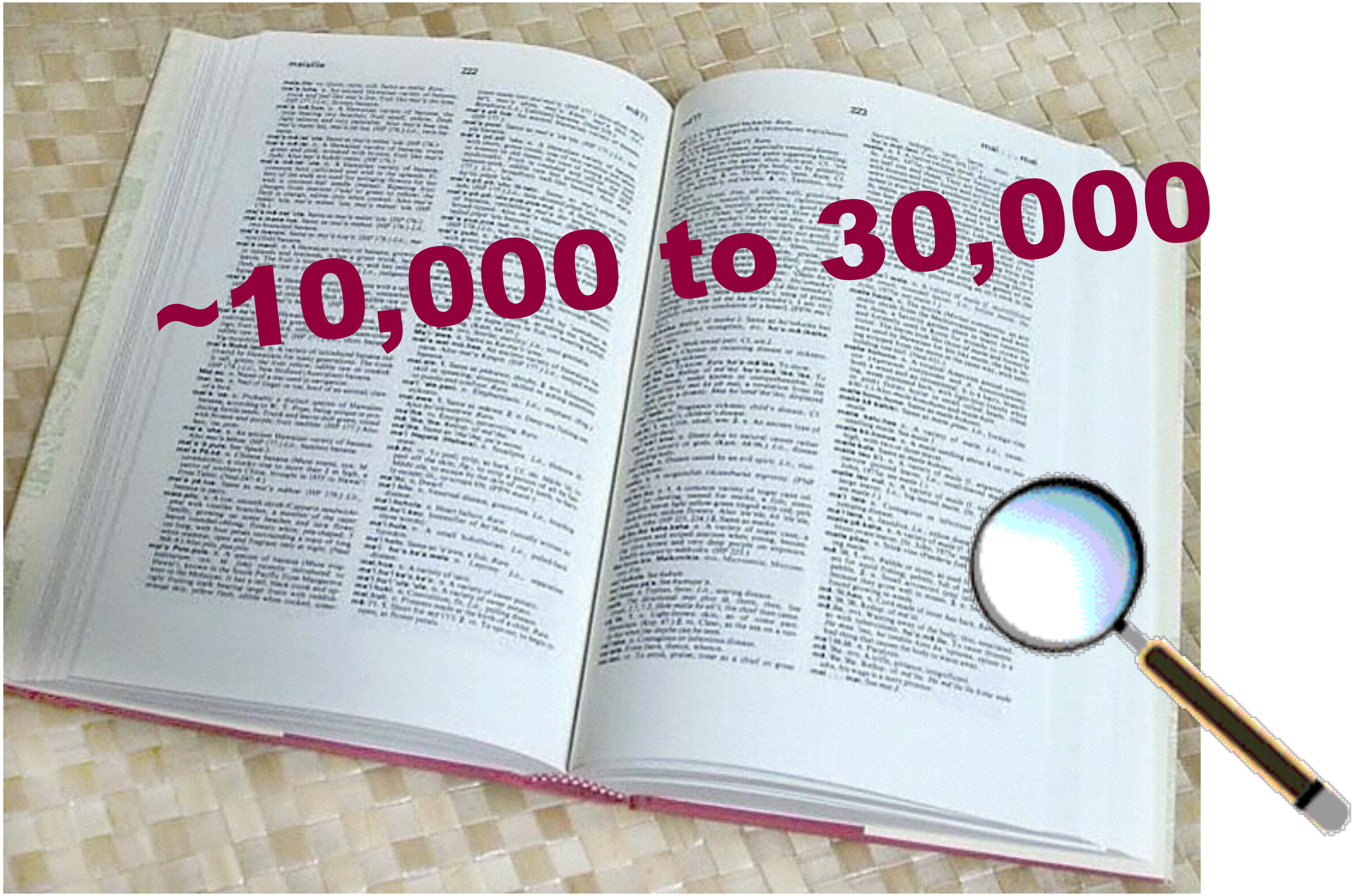


Image parsing

Image understanding?

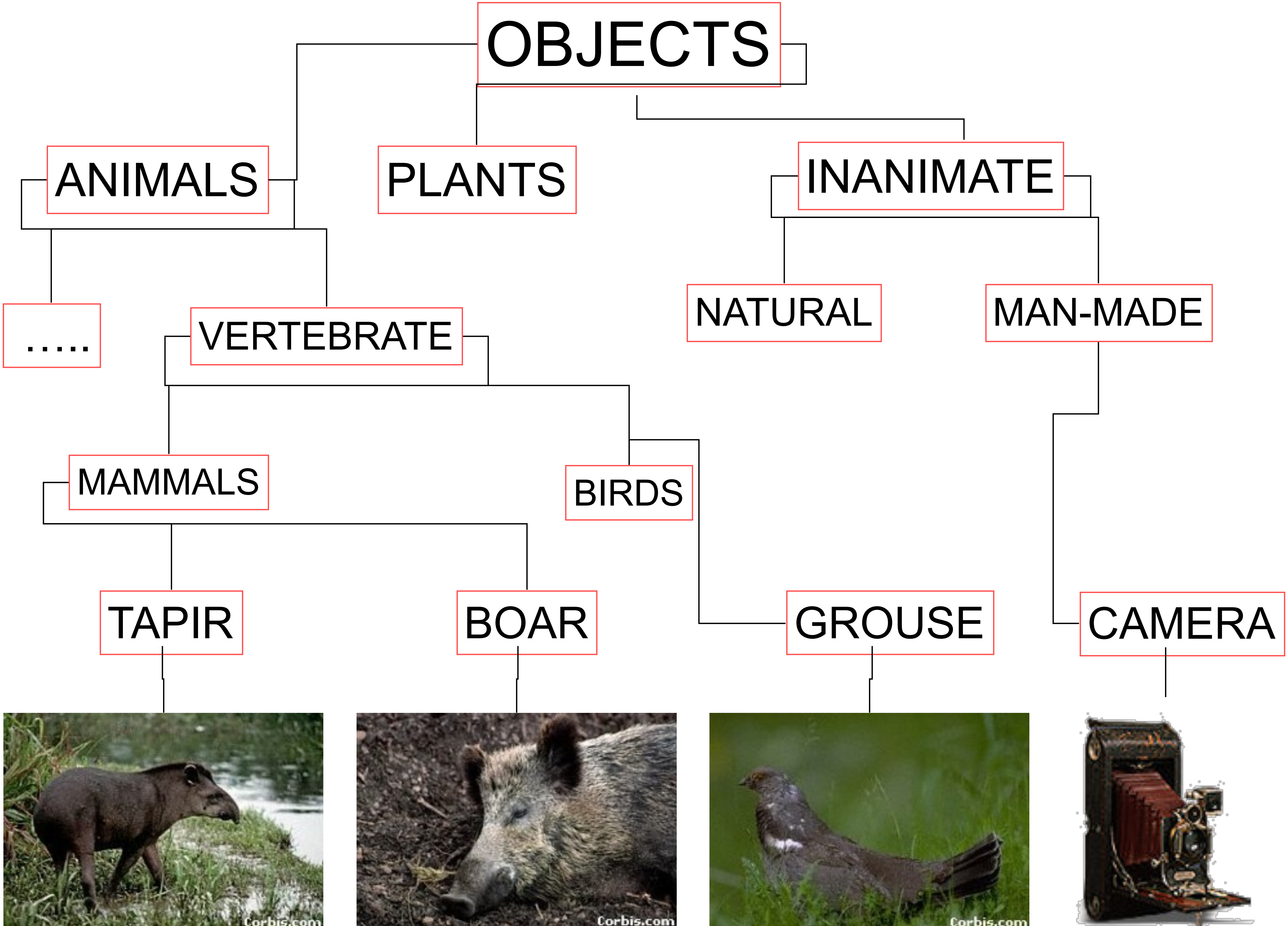
How many visual object categories are there?



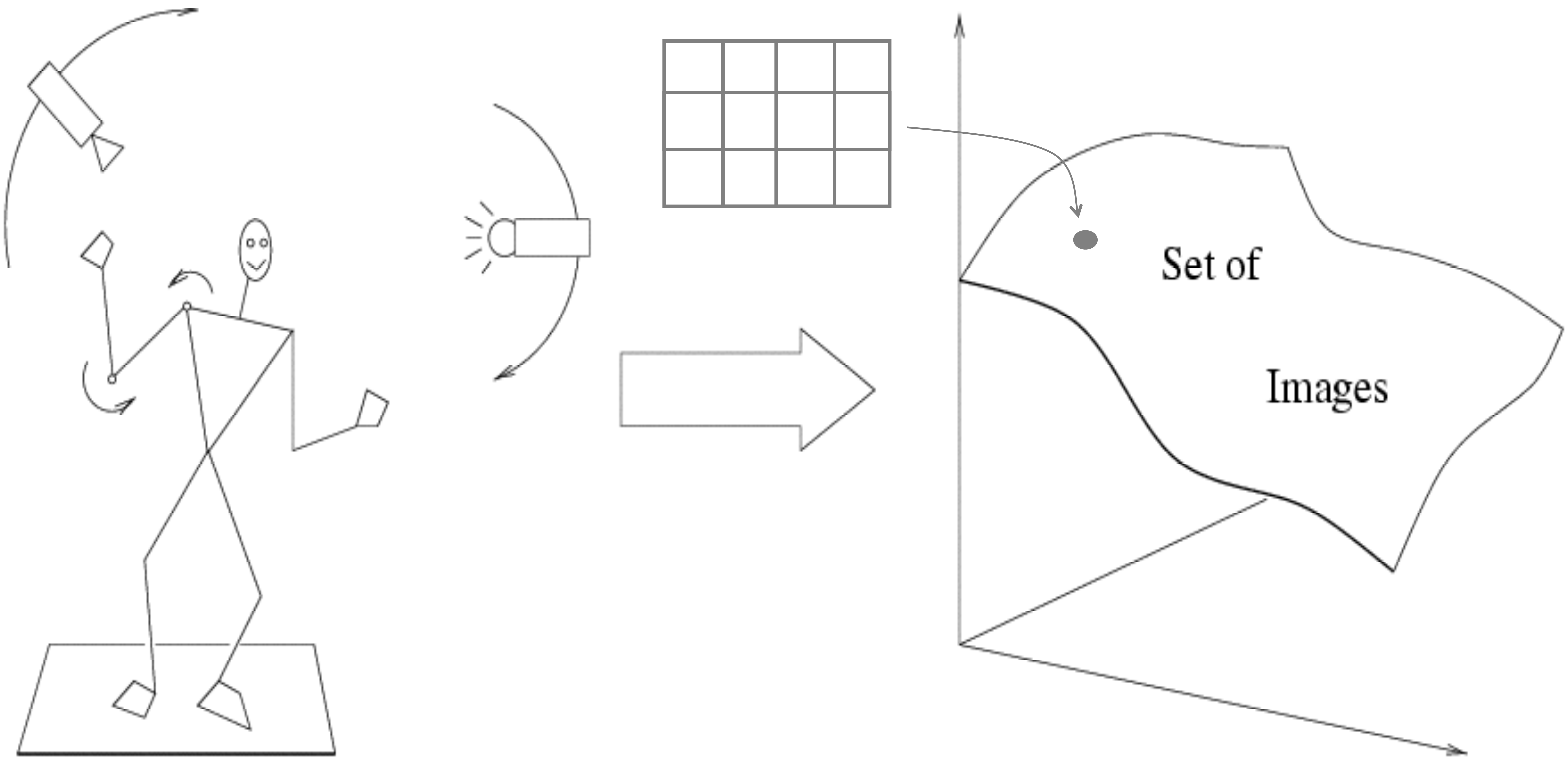
http://wexler.free.fr/library/files/biederman%20(1987)%20recognition-by-components.%20a%20theory%20of%20human%20image%20understanding.pdf

Biederman 1987


~10,000 to 30,000
S


OBJECTS
ANIMALS
PLANTS
INANIMATE
.....
VERTEBRATE
NATURAL
MAN-MADE
MAMMALS
BIRDS
TAPIR
BOAR
GROUSE
CAMERA
Corbis.com
Corbis.com
Corbis.com

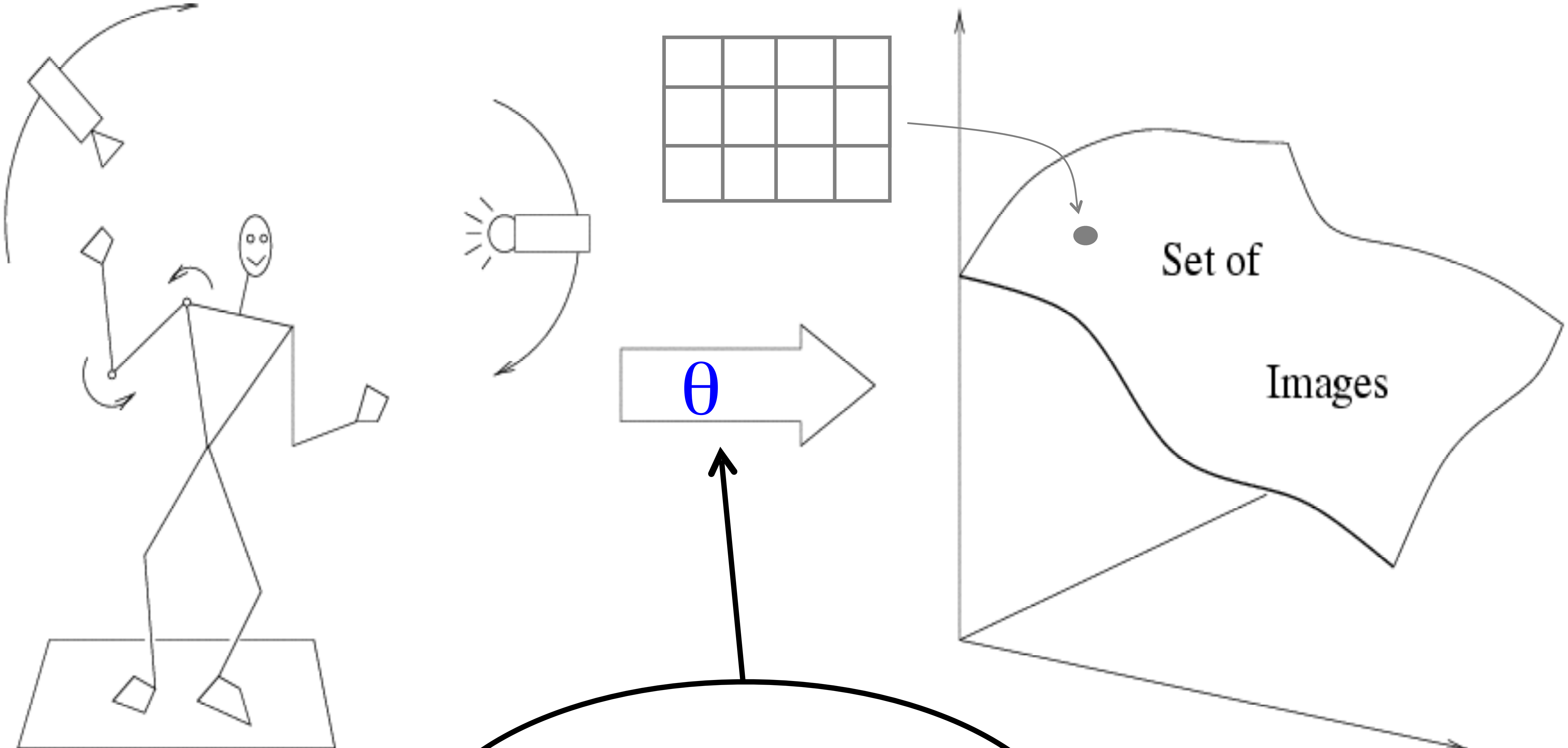
Recognition is all about modeling variability



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

History of ideas in recognition

1960s – early 1990s: the geometric era



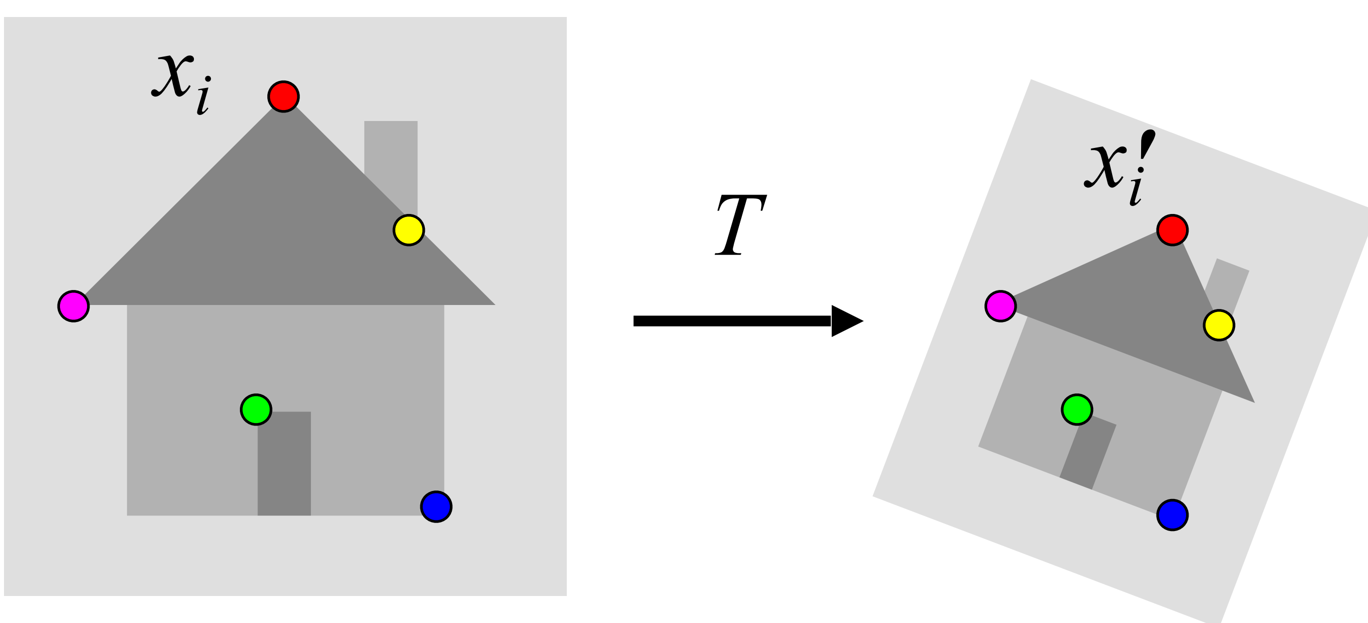
Variability: camera position **Alignment**

Shape: assumed known

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

Recall: Alignment

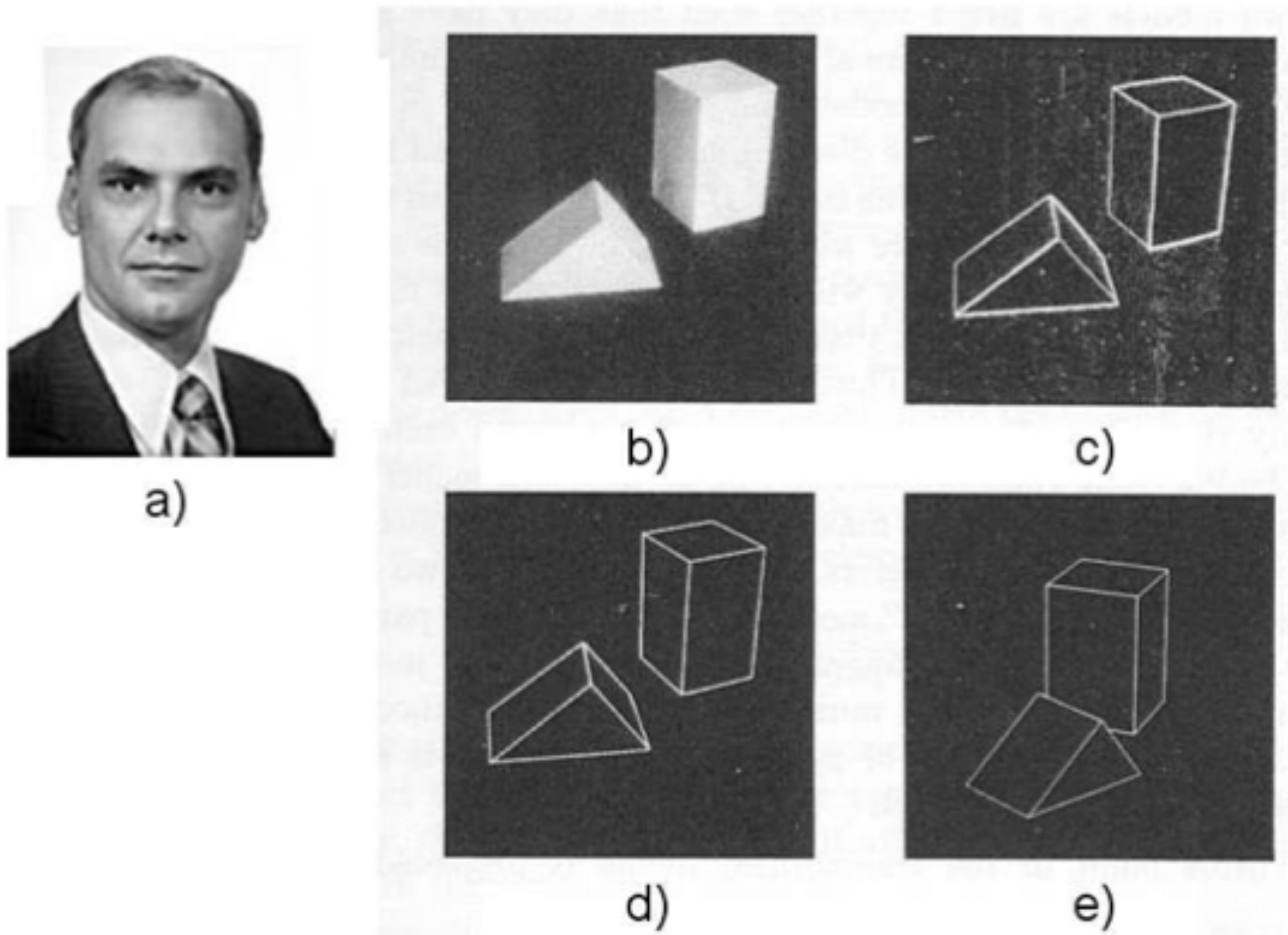
Alignment: fitting a model to a transformation between pairs of features (*matches*) in two images



Find transformation T that minimizes

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

Recognition as an alignment problem: Block world

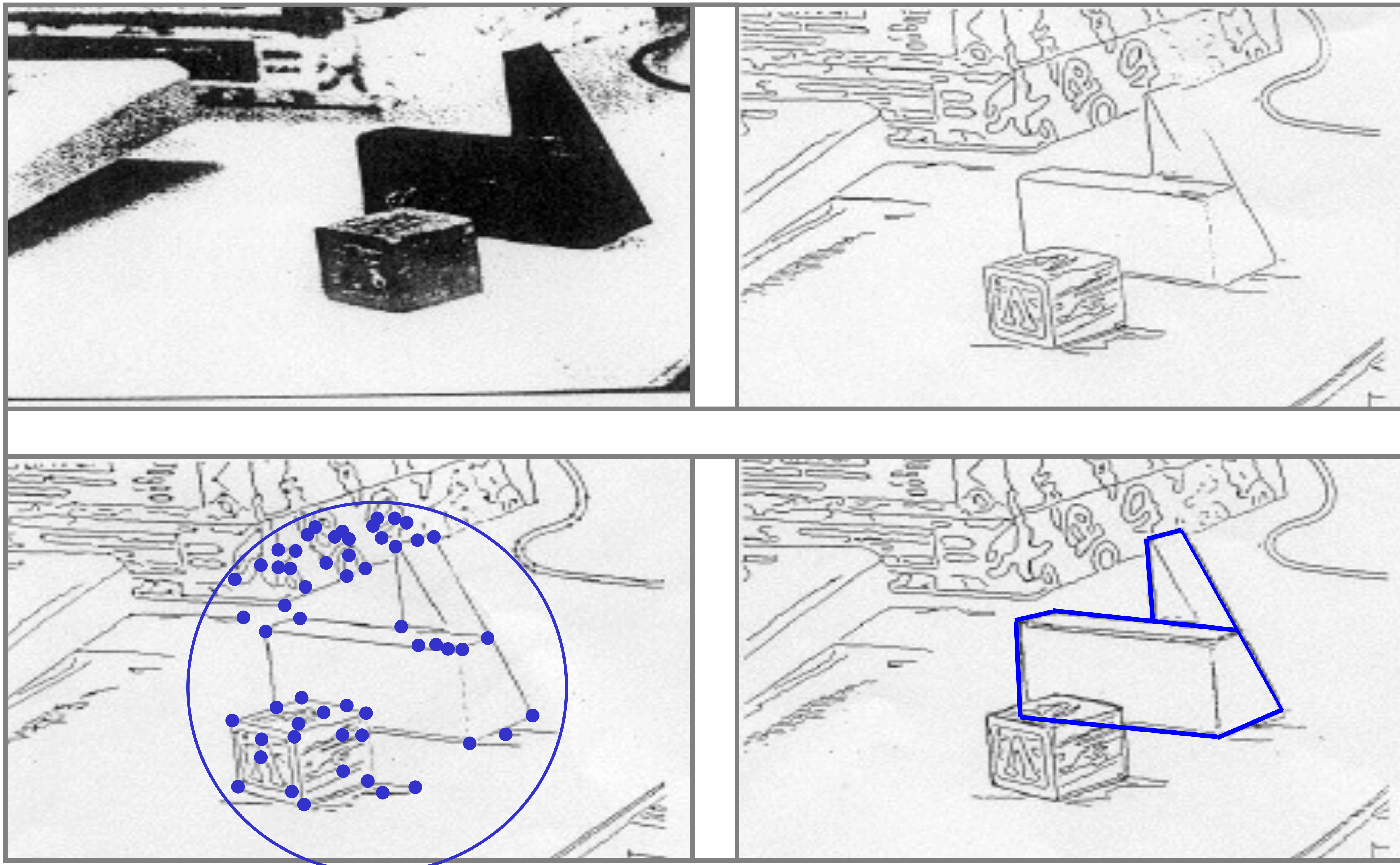


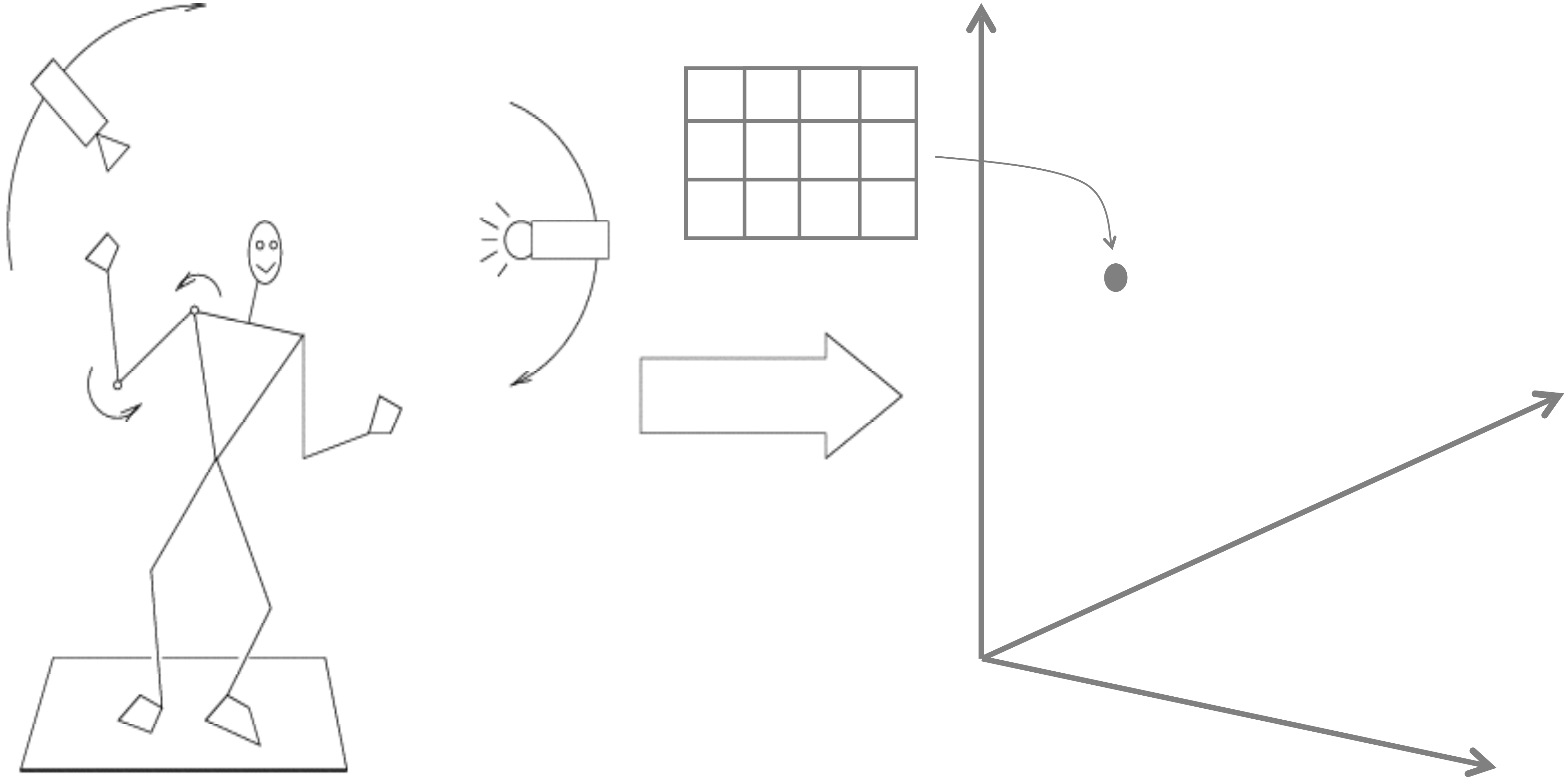
L. G. Roberts, [*Machine Perception of Three Dimensional Solids*](), Ph.D. thesis, MIT Department of Electrical Engineering, 1963.

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)





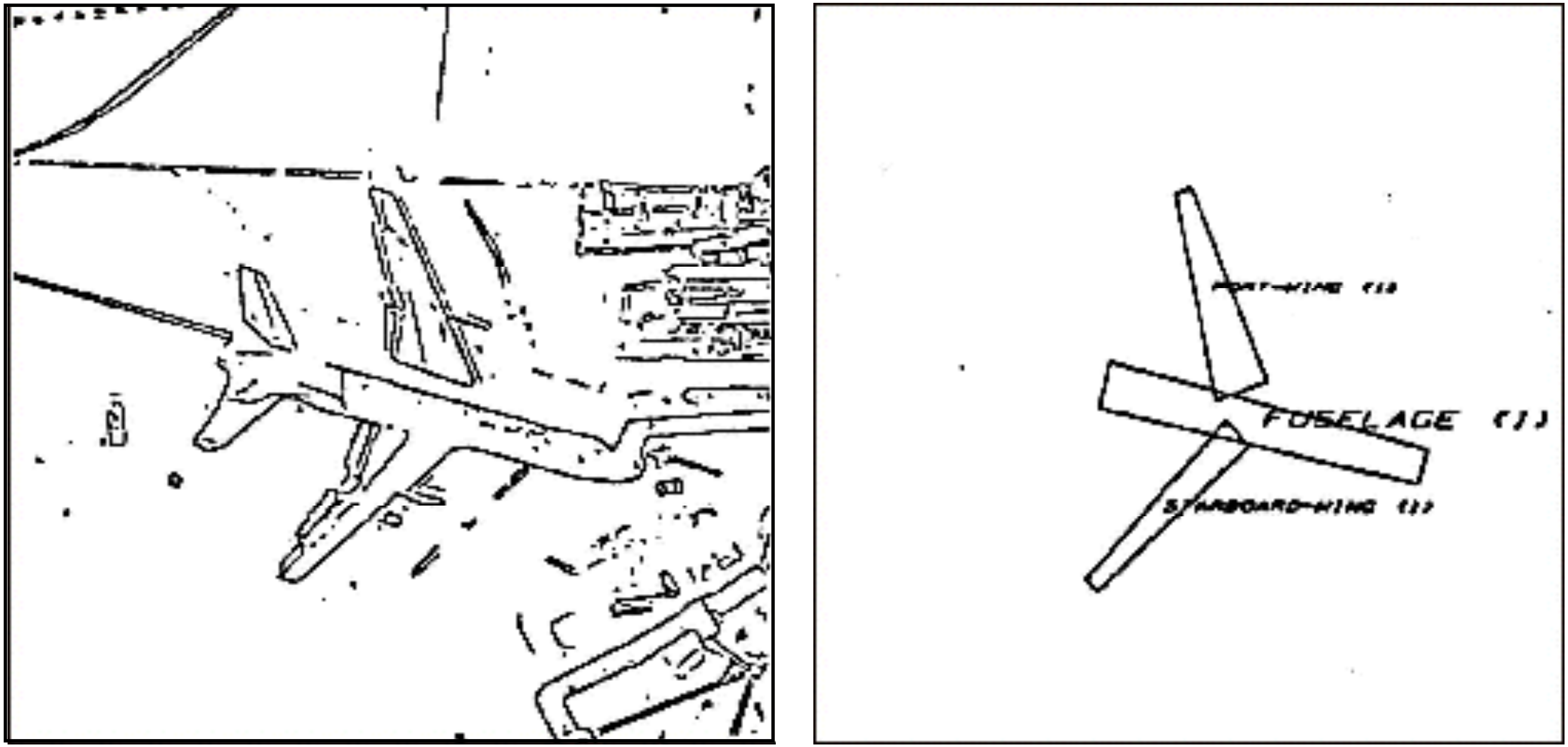
~~Variability~~

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)

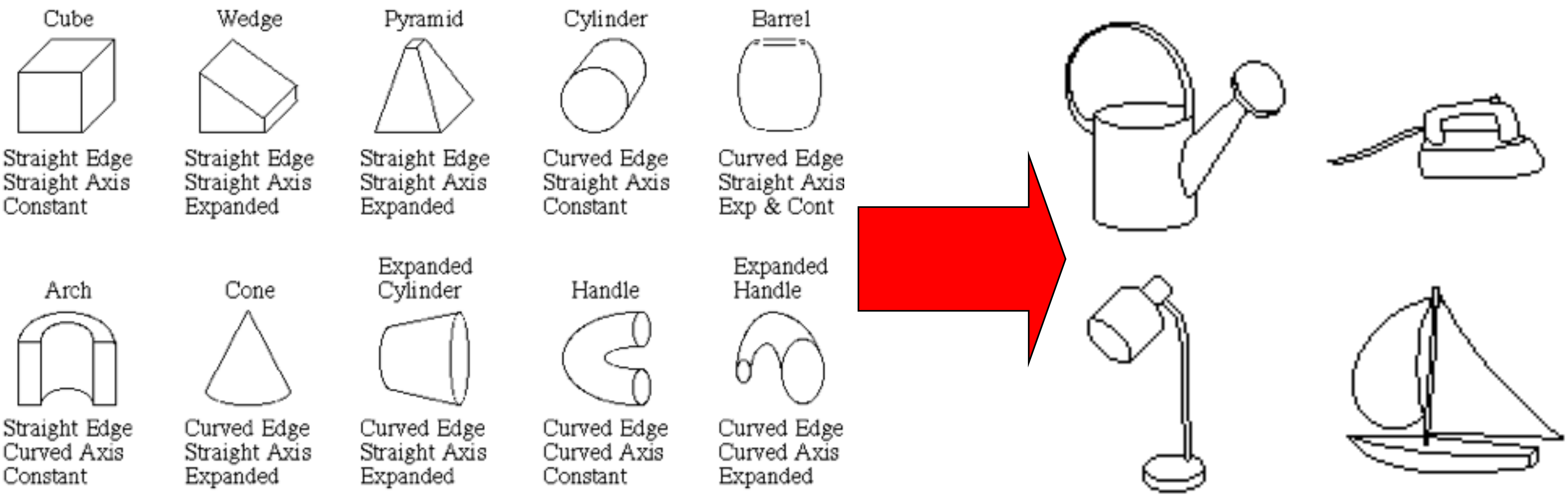
Binford (1971), Nevatia & Binford (1972), Marr & Nishihara (1978)

Recognition by components

Biederman (1987)

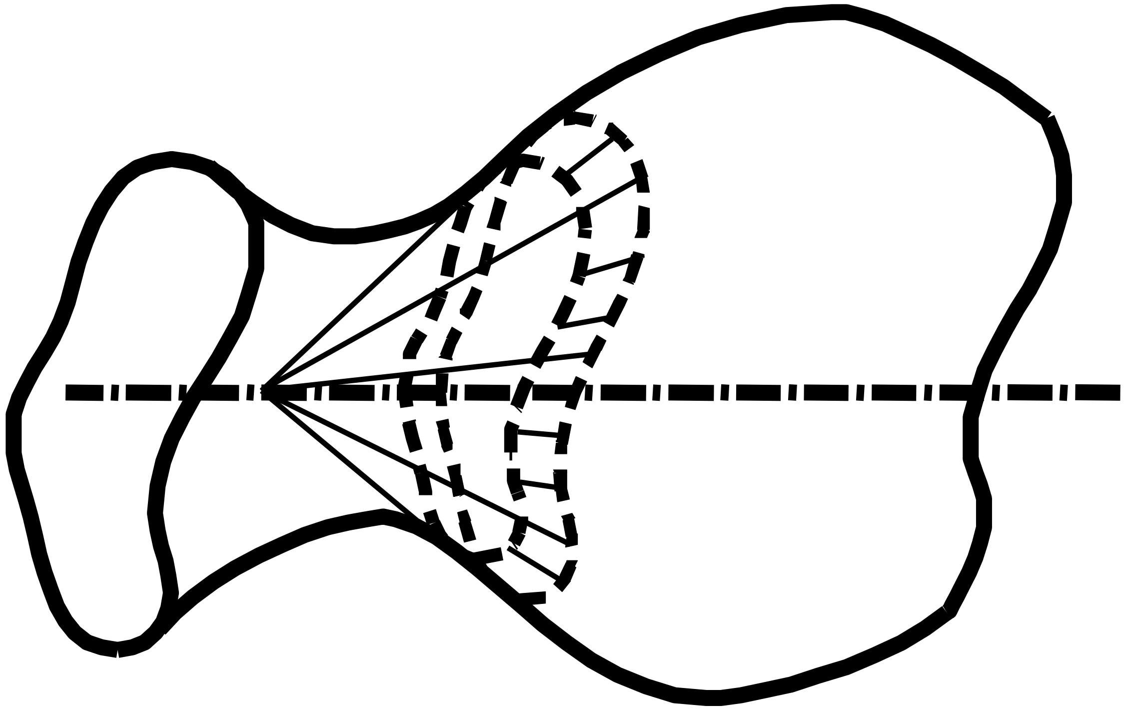
Primitives (*geons*)

Objects

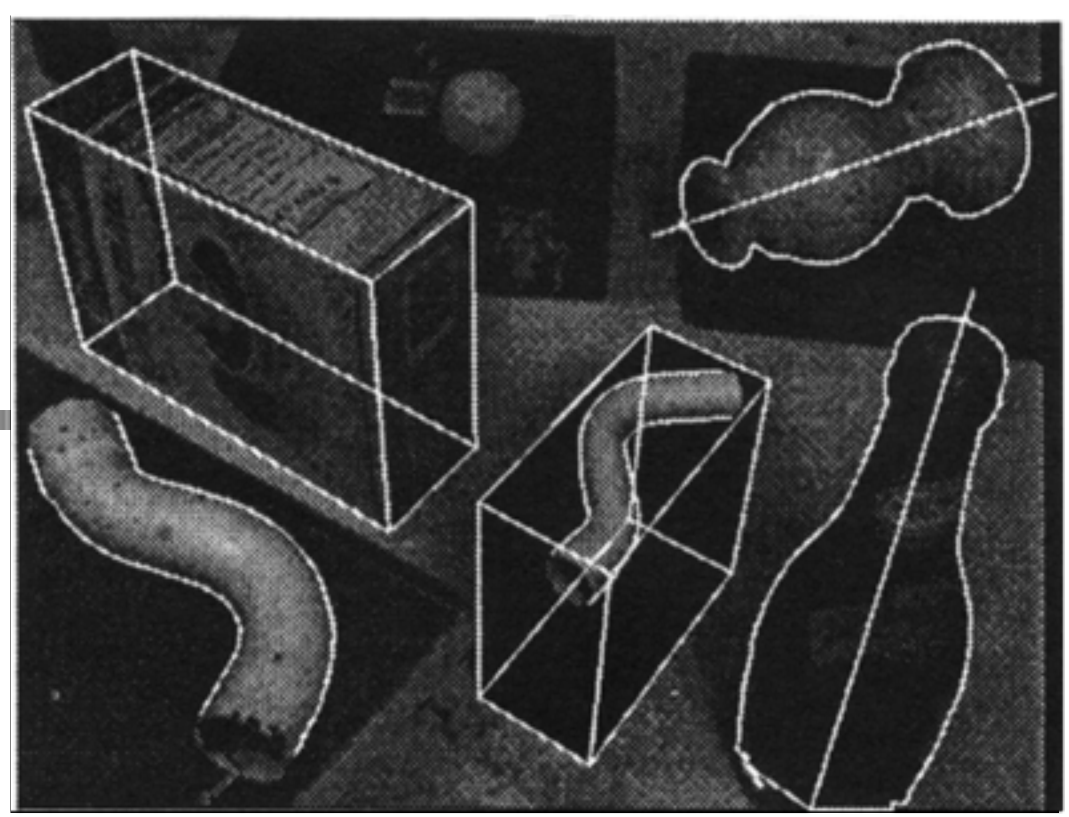


http://en.wikipedia.org/wiki/Recognition_by_Components_Theory

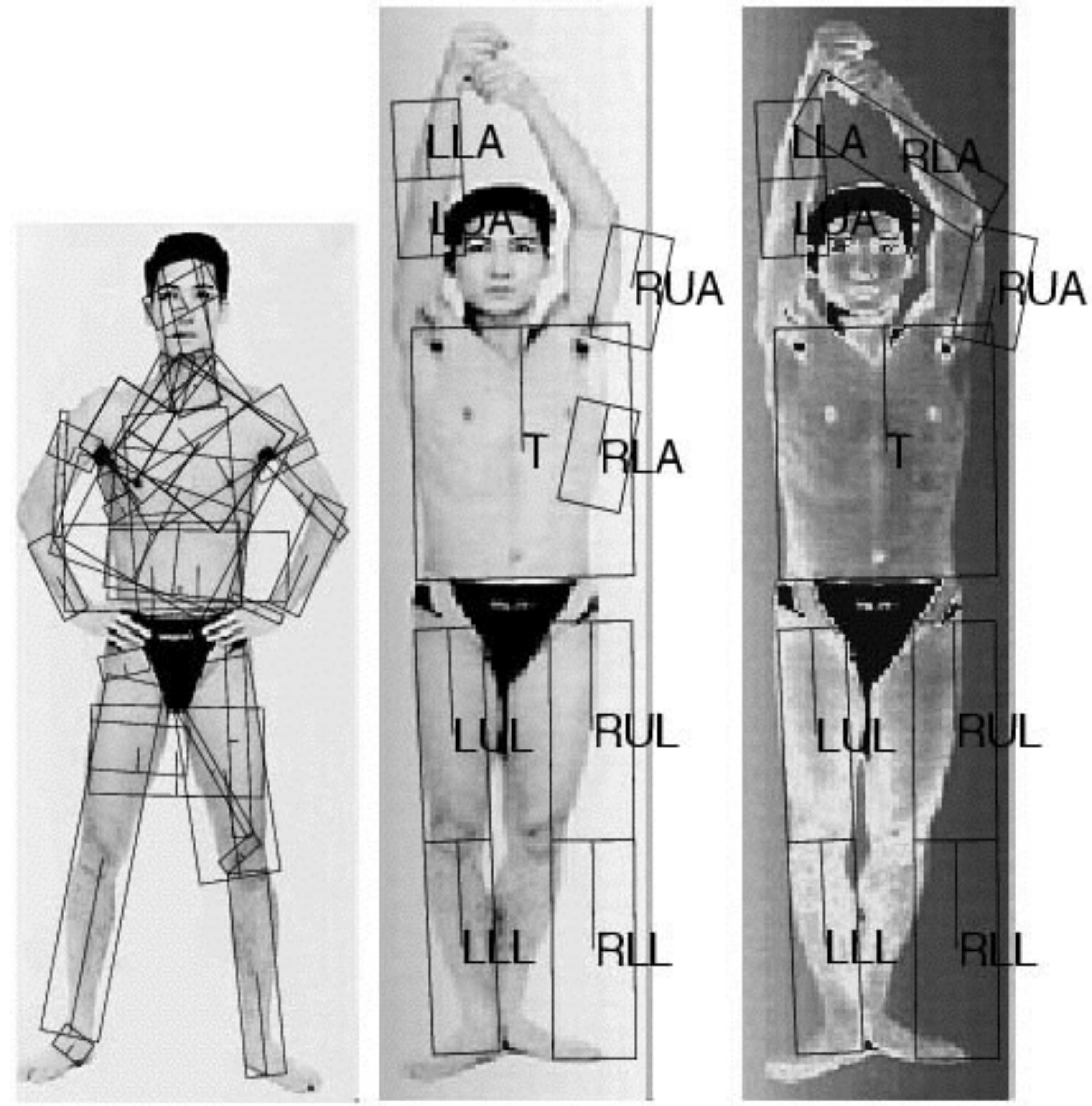
General shape primitives?



Generalized cylinders
Ponce et al. (1989)



Zisserman et al. (1995)

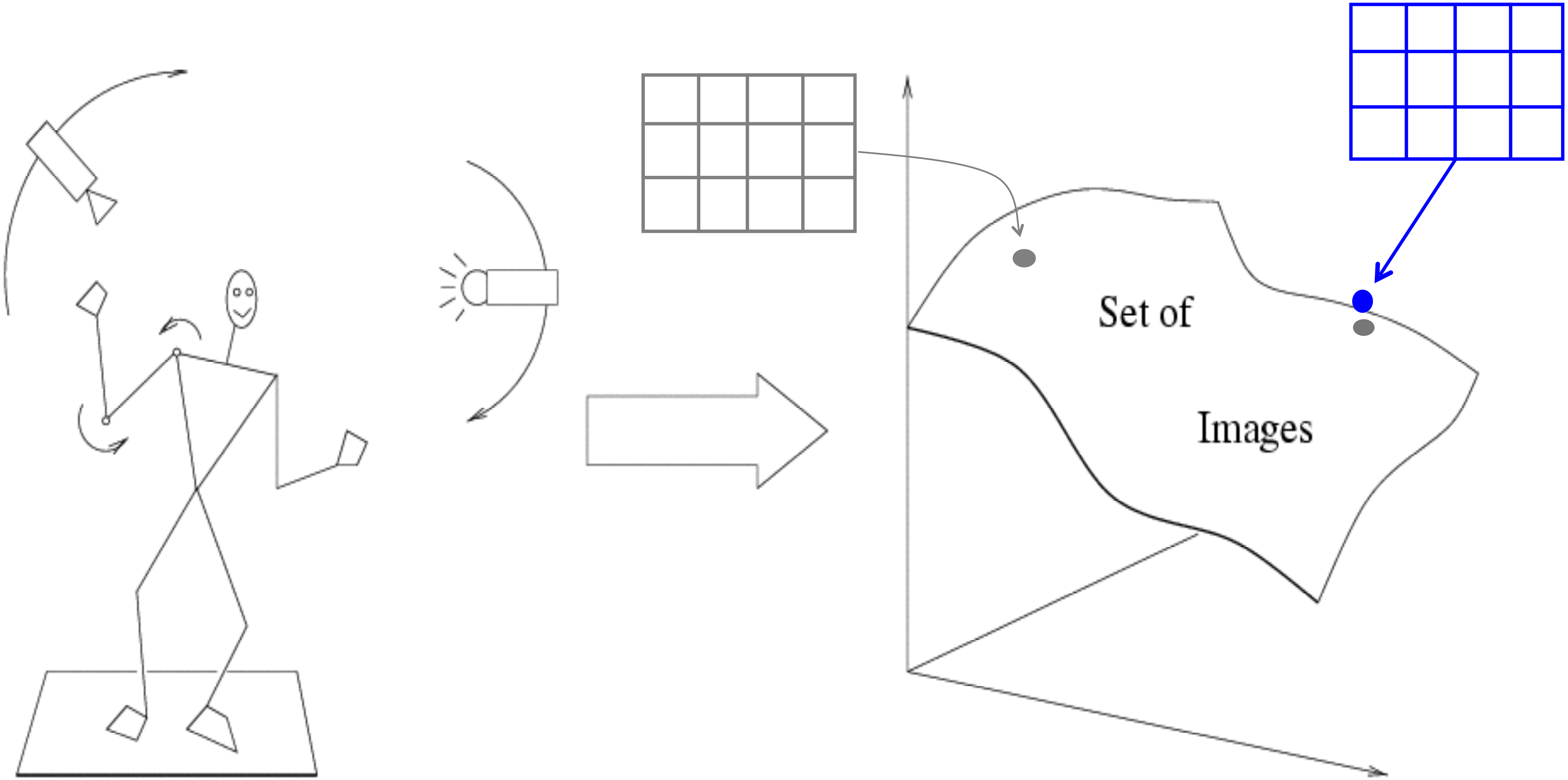


Forsyth (2000)

History of ideas in recognition

1960s – early 1990s: the geometric era

1990s: appearance-based models

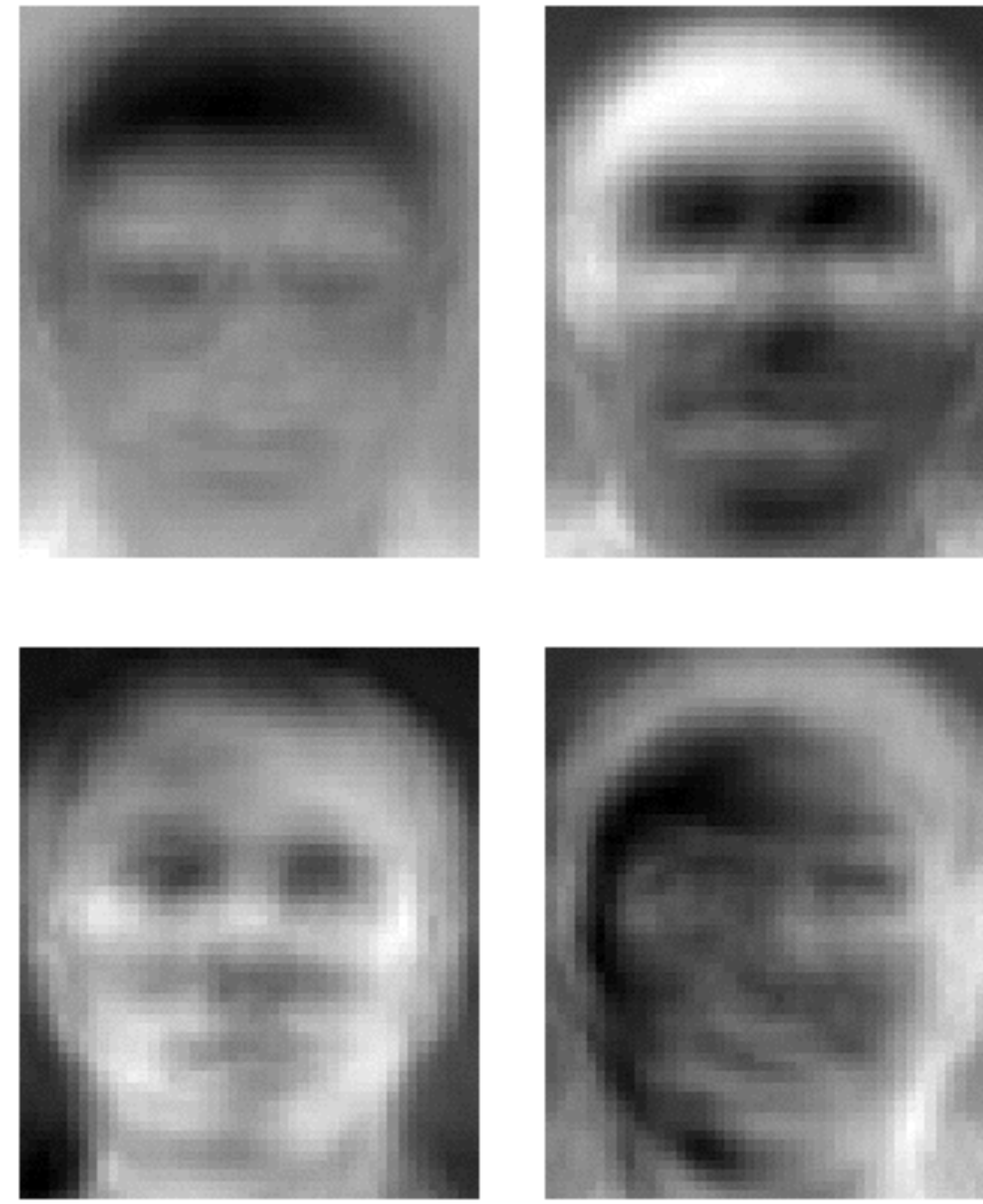


Empirical models of image variability

Appearance-based techniques

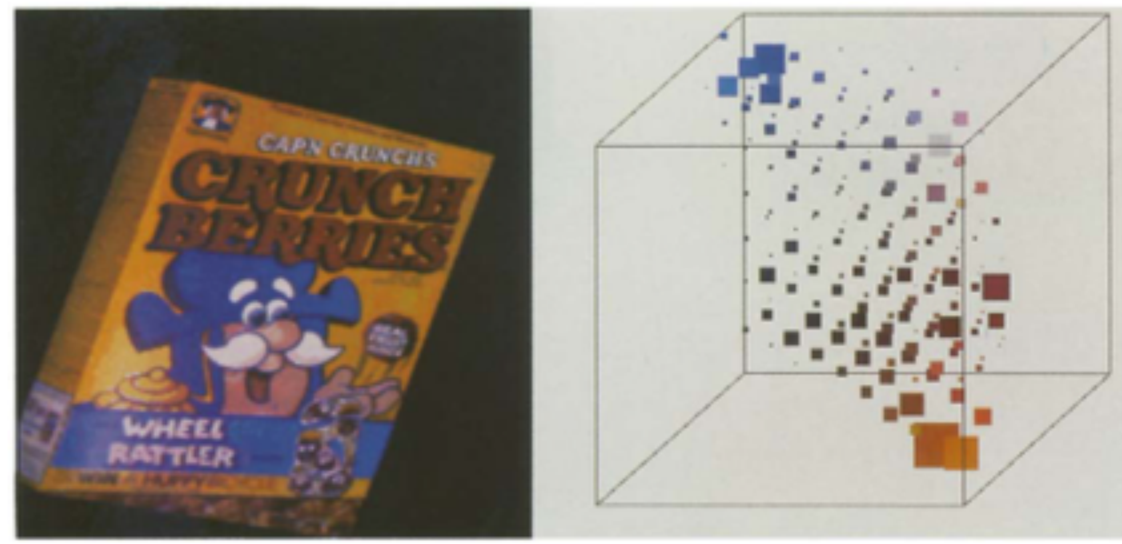
Turk & Pentland (1991); Murase & Nayar (1995); etc.

Eigenfaces (Turk & Pentland, 1991)



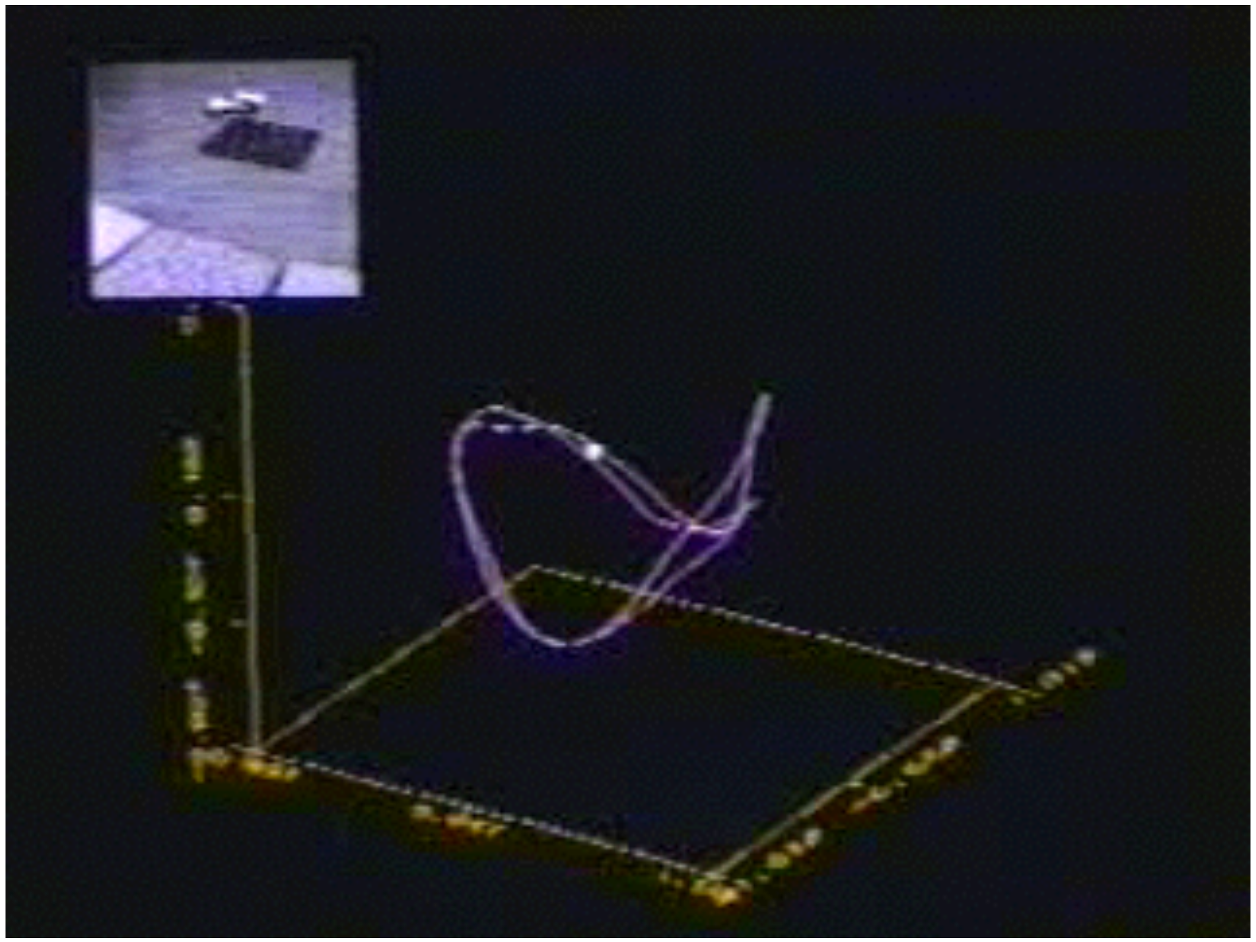
Experimental	Correct/Unknown Recognition Percentage		
Condition	Lighting	Orientation	Scale
Forced classification	96/0	85/0	64/0
Forced 100% accuracy	100/19	100/39	100/60
Forced 20% unknown rate	100/20	94/20	74/20

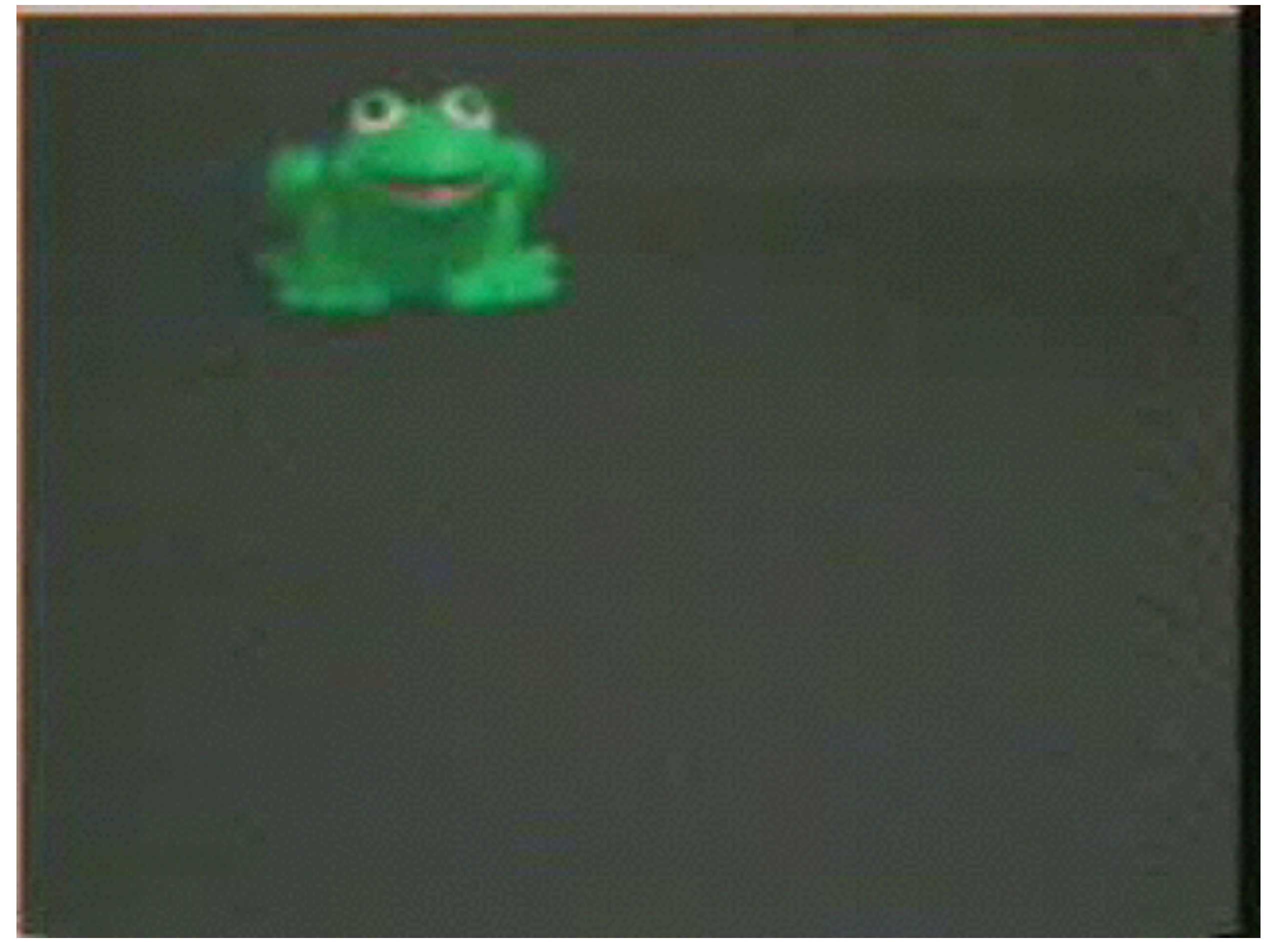
Color Histograms



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

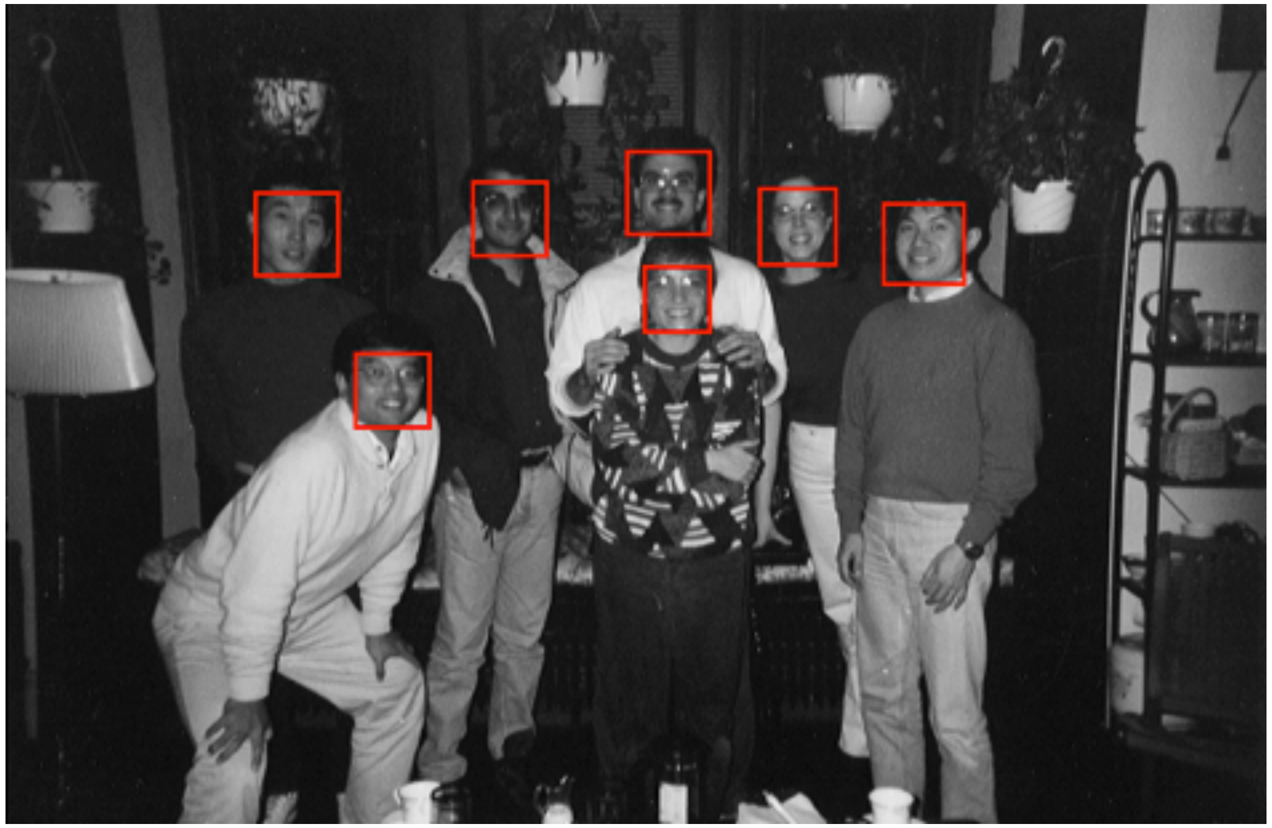
1960s – early 1990s: the geometric era

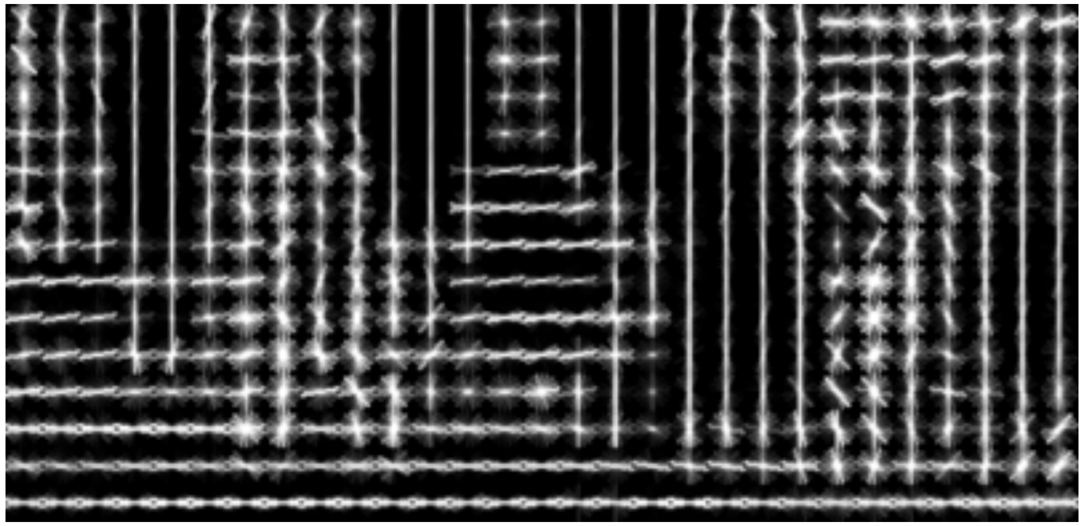
1990s: appearance-based models

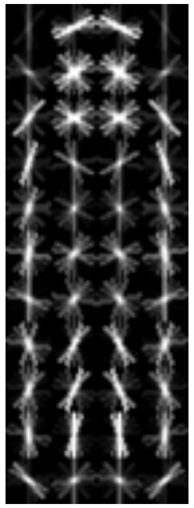
1990s – present: sliding window approaches

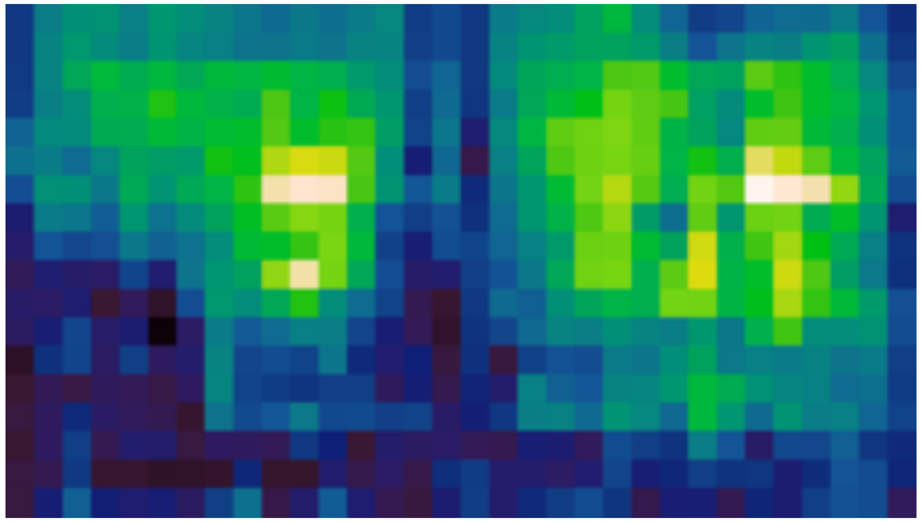
Sliding window approaches

Sliding window approaches









History of ideas in recognition

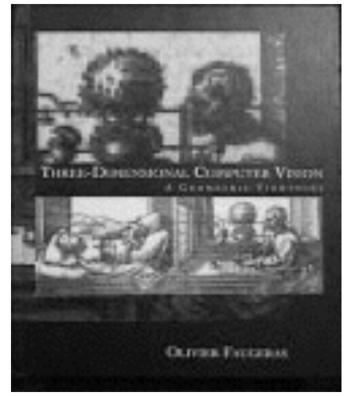
1960s – early 1990s: the geometric era

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

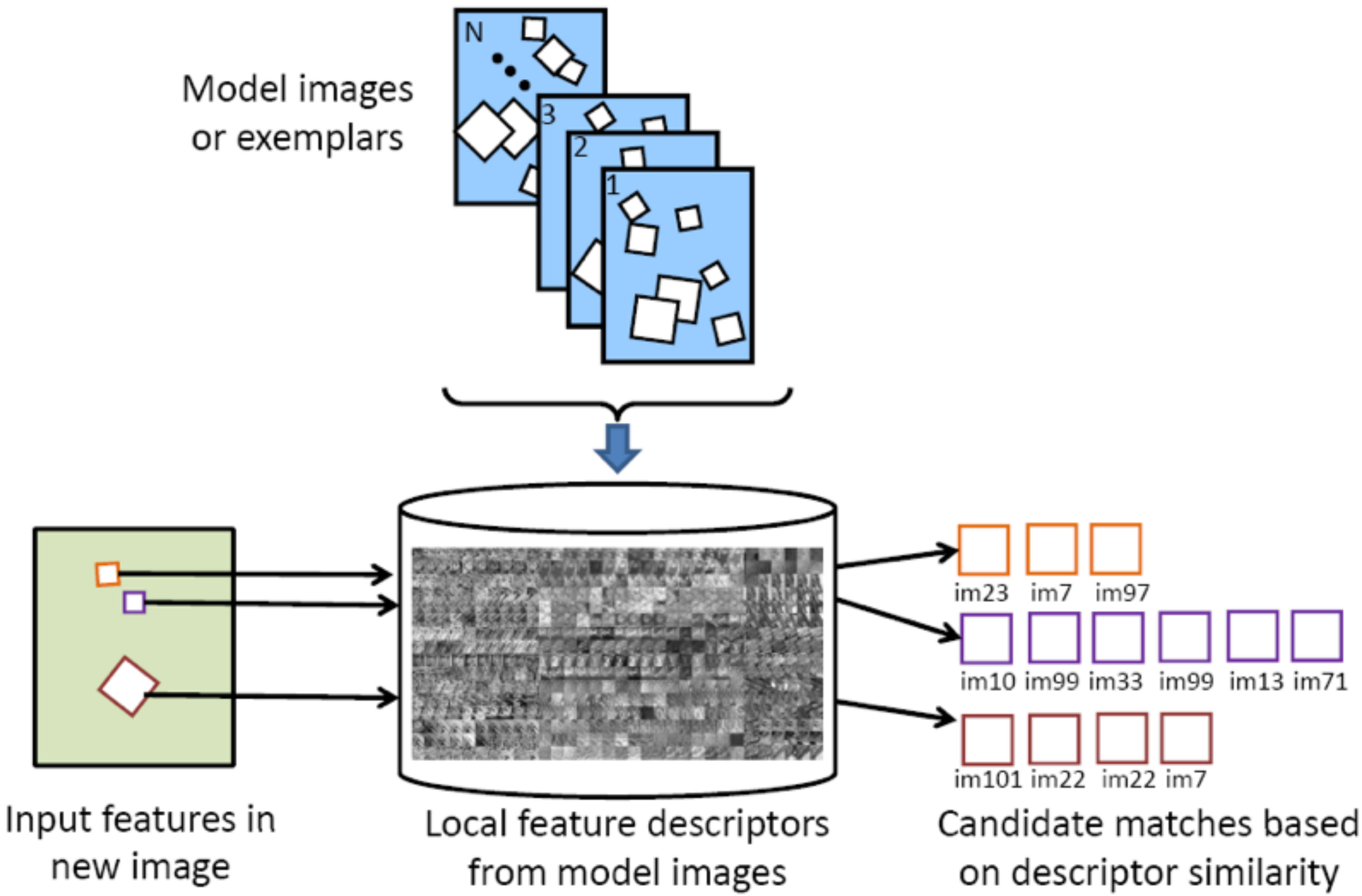


Image credit: K. Grauman and B. Leibe

Large-scale image search

Combining local features, indexing, and spatial constraints

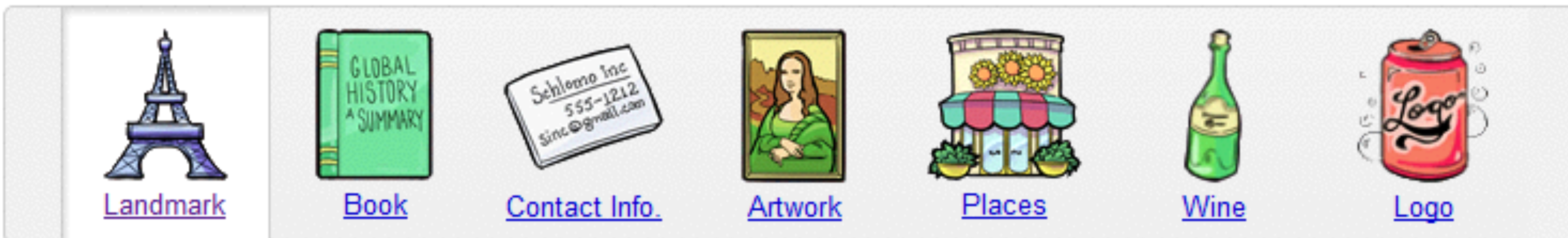
Philbin et al. ‘07

Large-scale image search

Combining local features, indexing, and spatial constraints

Google Goggles in Action

Click the icons below to see the different ways Google Goggles can be used.



History of ideas in recognition

1960s – early 1990s: the geometric era

1990s: appearance-based models

1990s – present: sliding window approaches

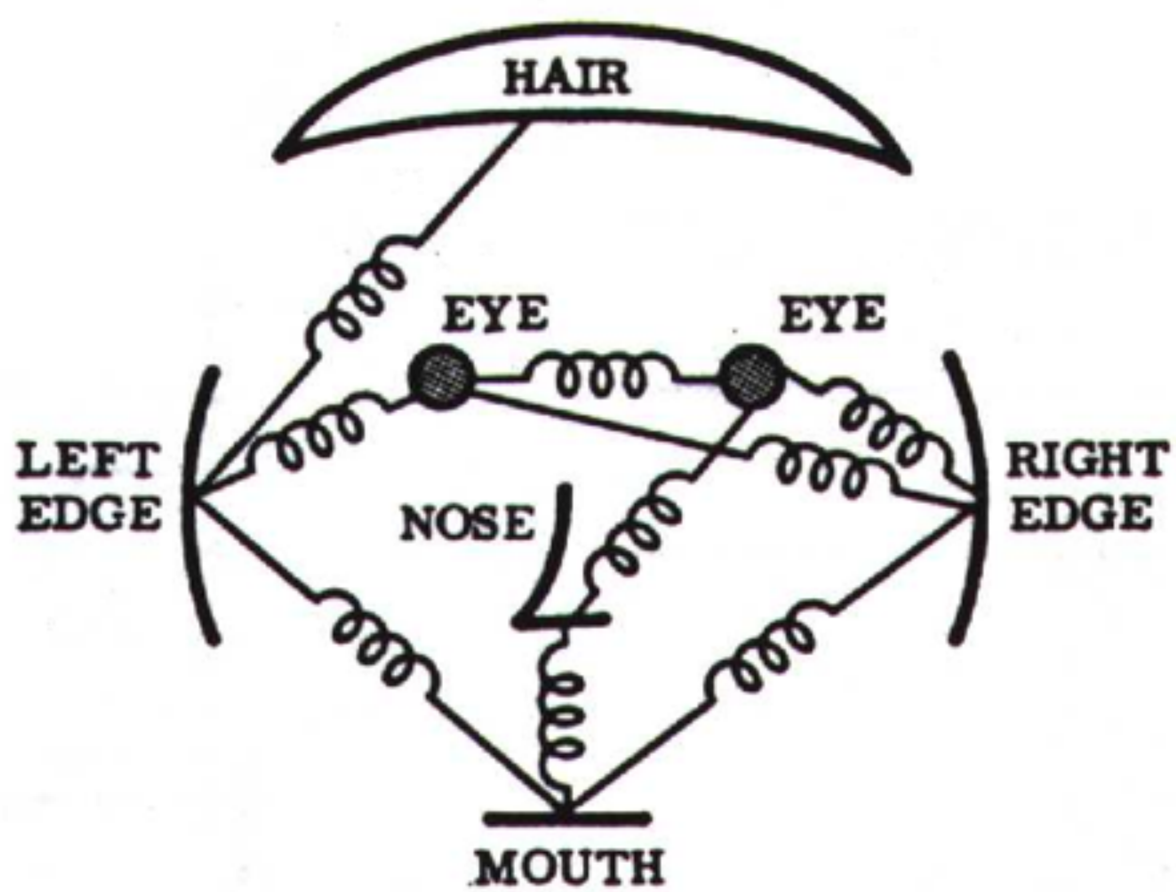
Late 1990s: local features

Early 2000s: parts-and-shape models

Parts-and-shape models

Model:

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



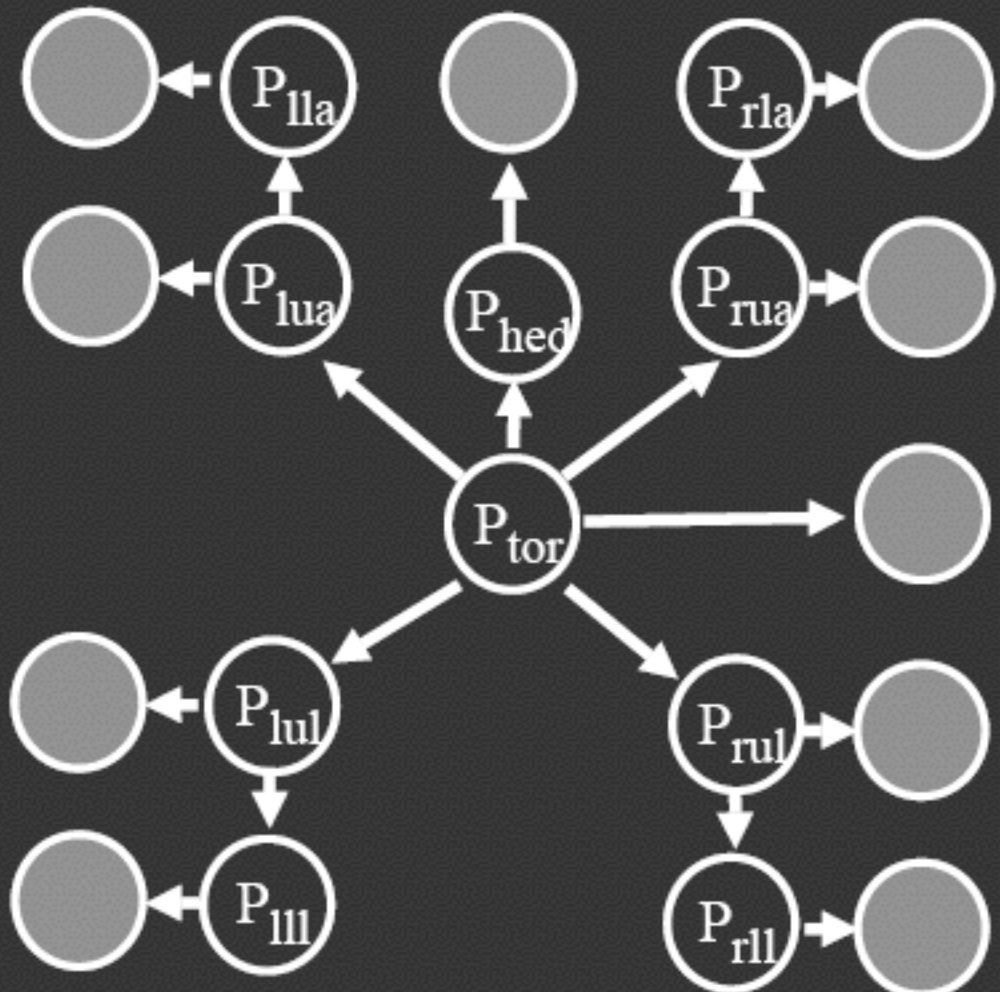
Fischler & Elschlager 73

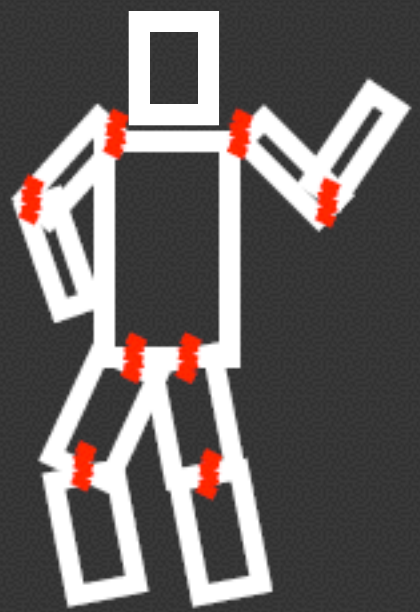
Constellation models

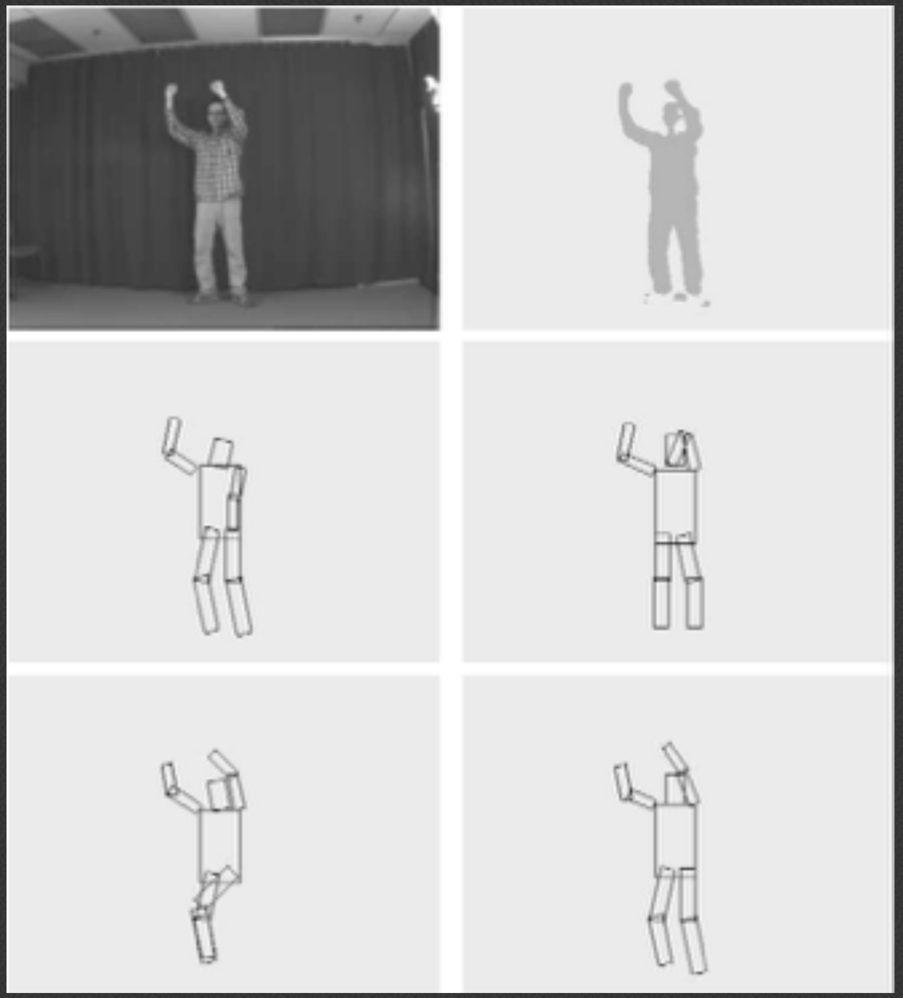
Weber, Welling & Perona (2000), Fergus, Perona & Zisserman (2003)

Pictorial structure model

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







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

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

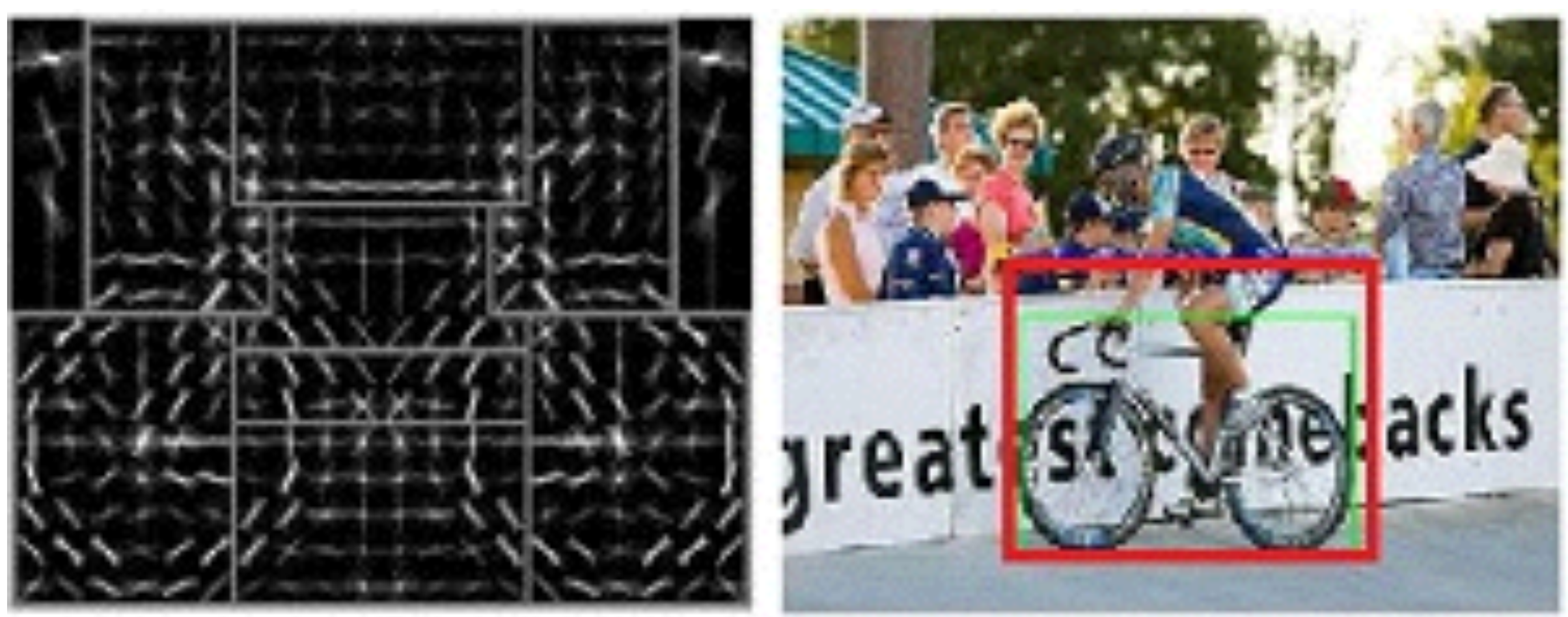
All of these are treated as being the same

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

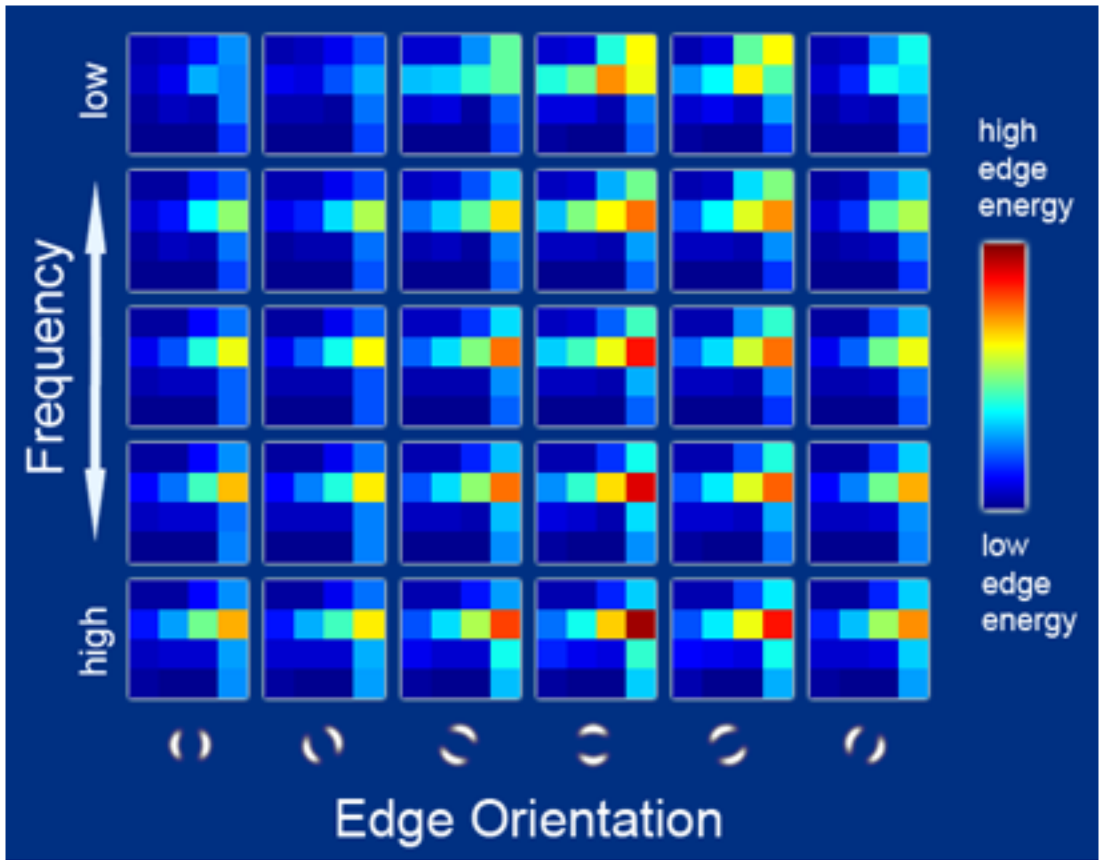
Early 2000s: parts-and-shape models

Mid-2000s: bags of features

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

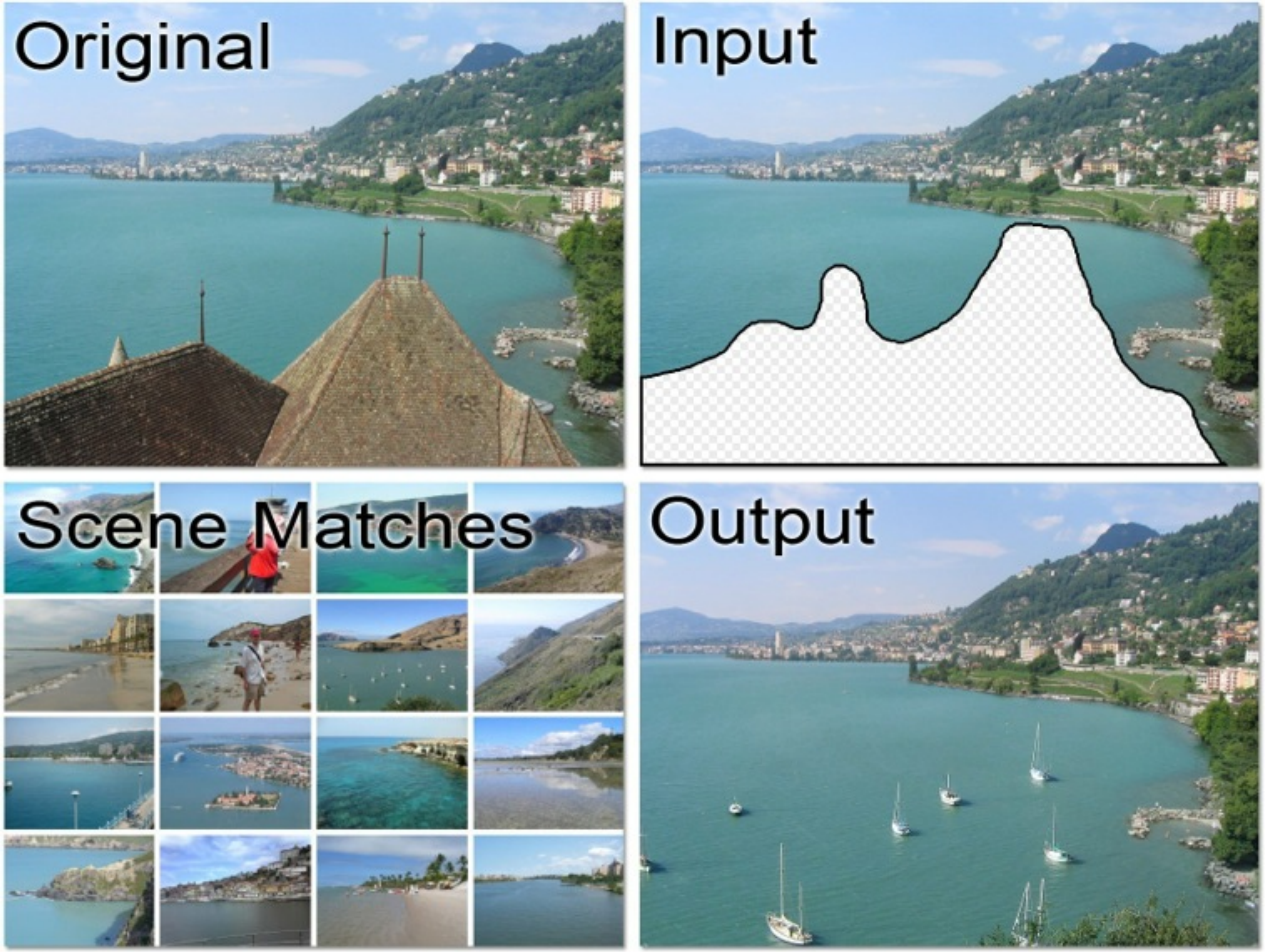
Global appearance models revisited

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



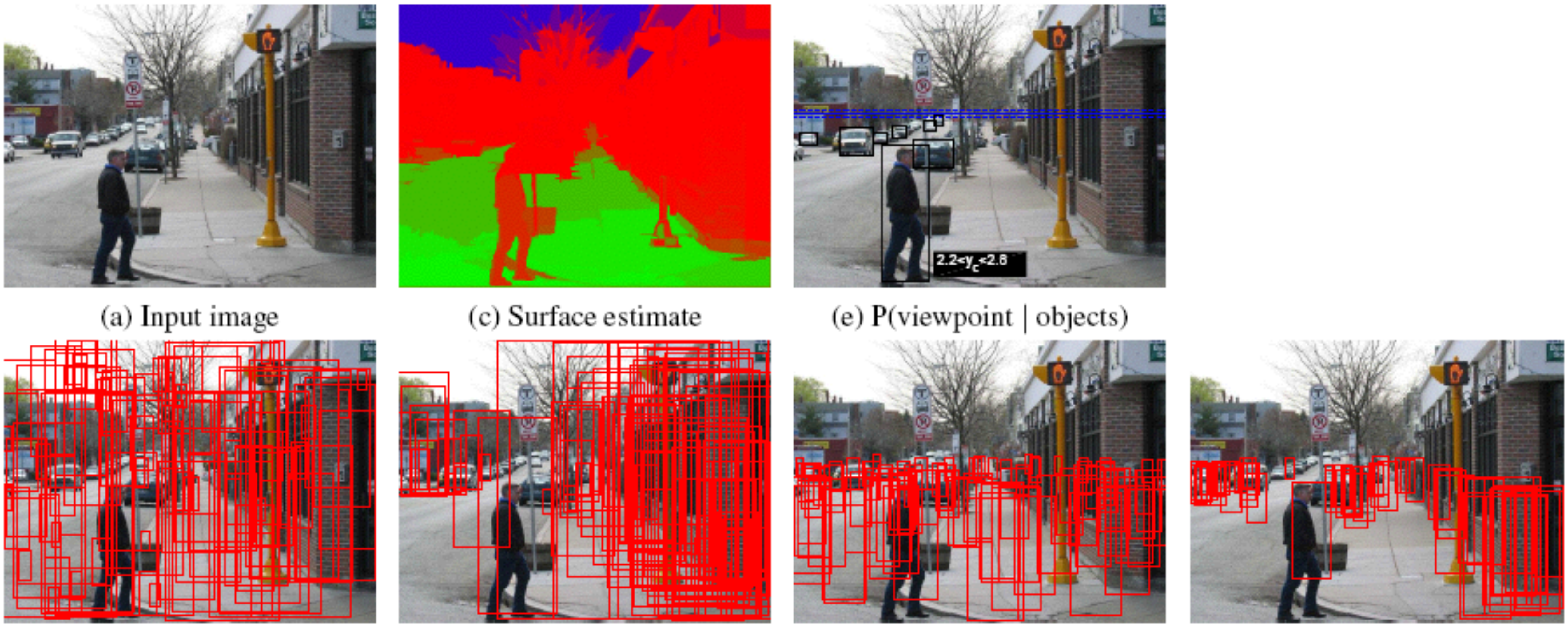
http://people.csail.mit.edu/torralba/code/spatialenvelope/

New applications in graphics



J. Hays and A. Efros, [Scene Completion using Millions of Photographs](), SIGGRAPH 2007

Geometric context



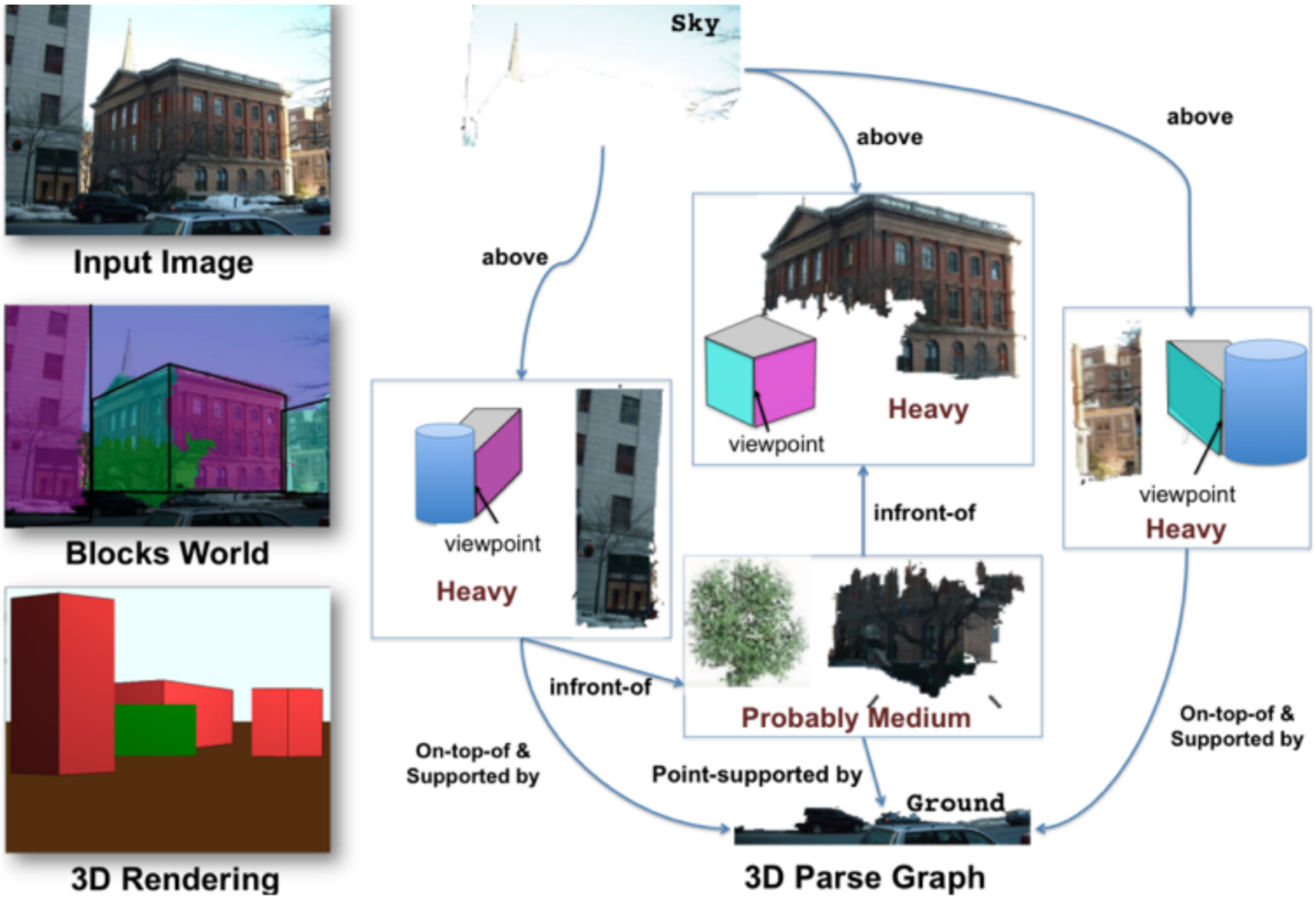
(a) Input image (c) Surface estimate (e) P(viewpoint | objects)
(b) P(person) = uniform (d) P(person | geometry) (f) P(person | viewpoint) (g) P(person|viewpoint,geometry)

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

Geometry and recognition

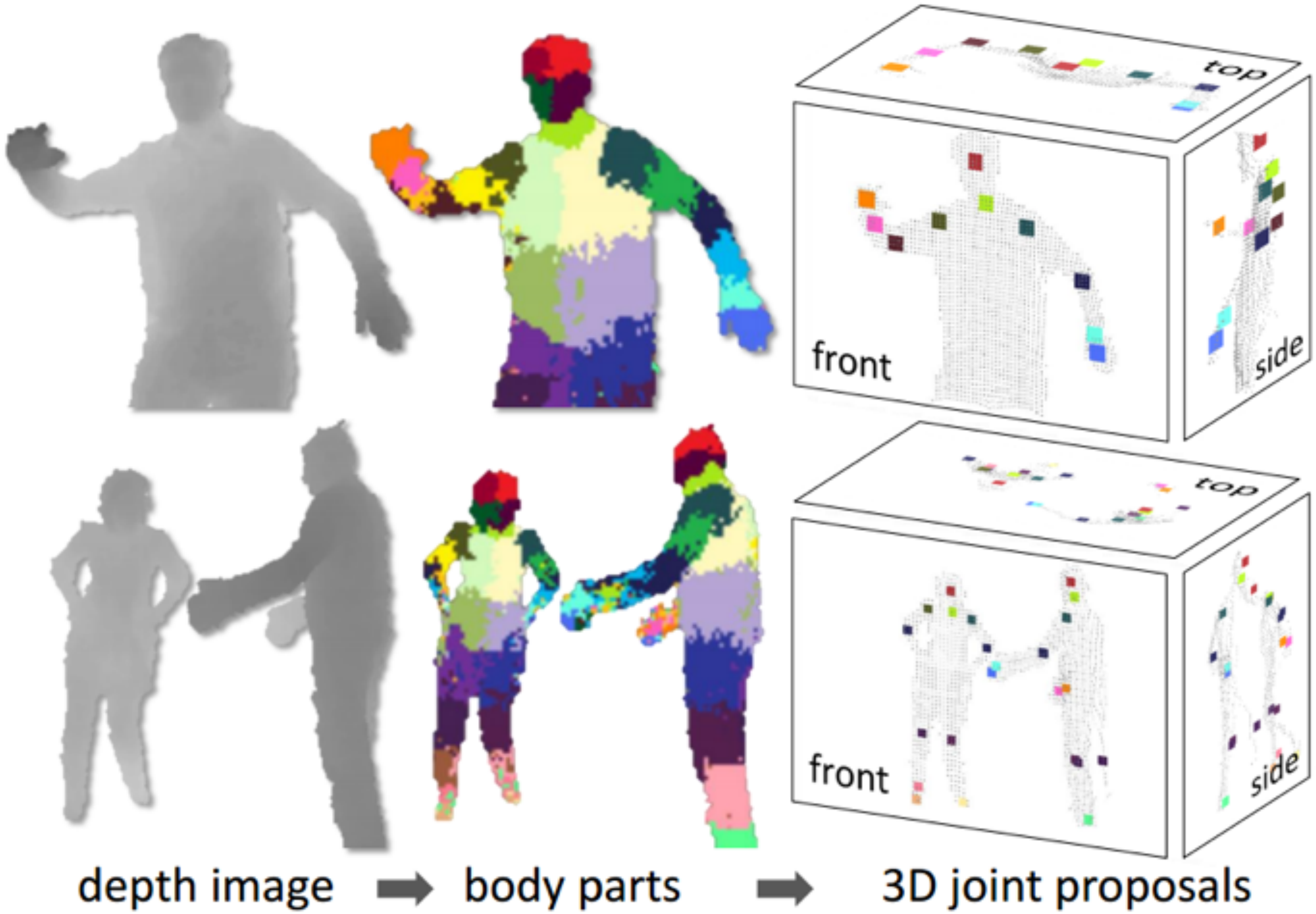
V. Hedau, D. Hoiem, and D. Forsyth, **[Recovering the Spatial Layout of Cluttered Rooms]()**, ICCV 2009.

Geometry and recognition



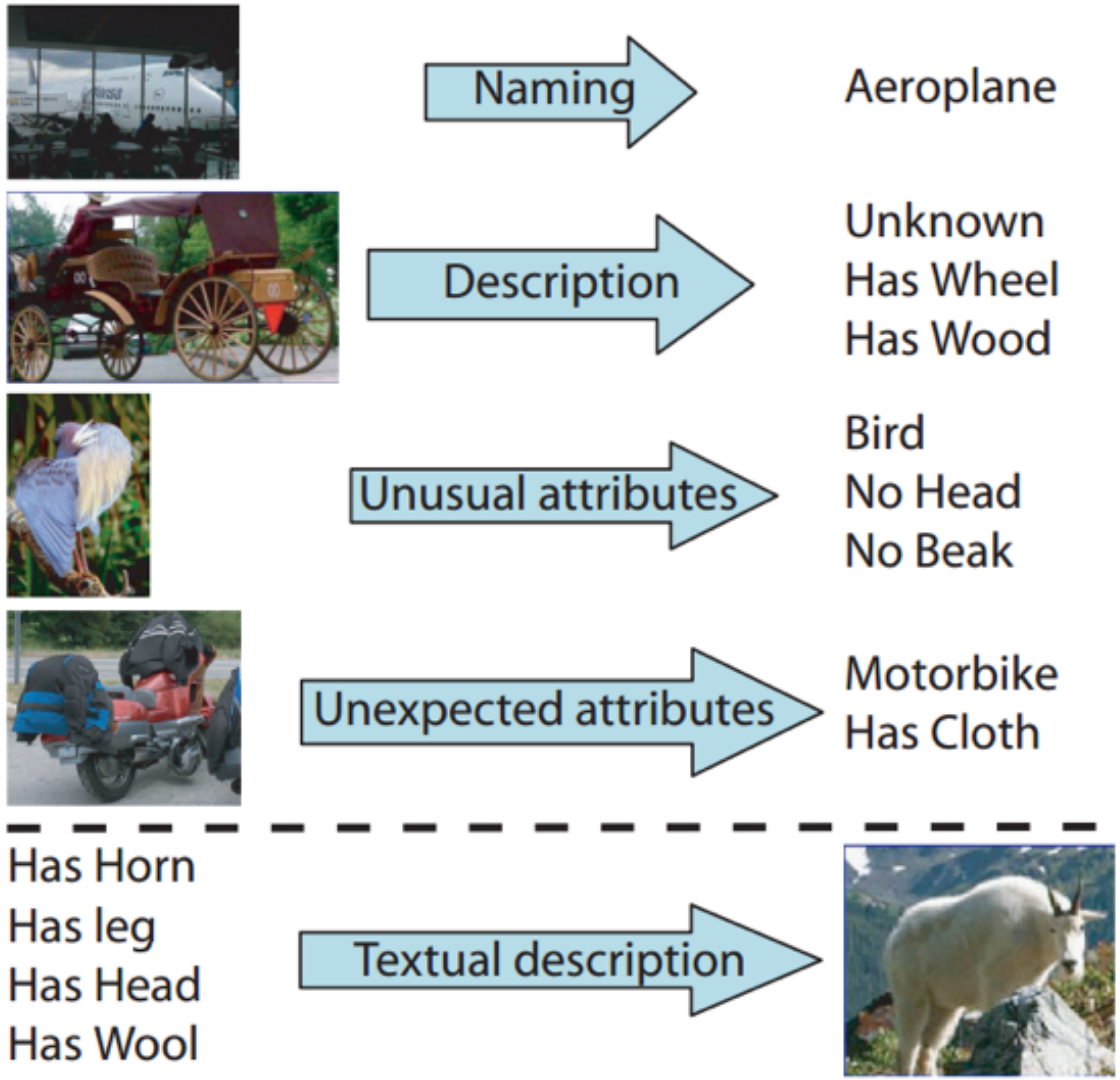
A. Gupta, A. Efros and M. Hebert, **Blocks World Revisited: Image Understanding Using Qualitative Geometry and Mechanics**, ECCV 2010

Recognition from RGBD Images



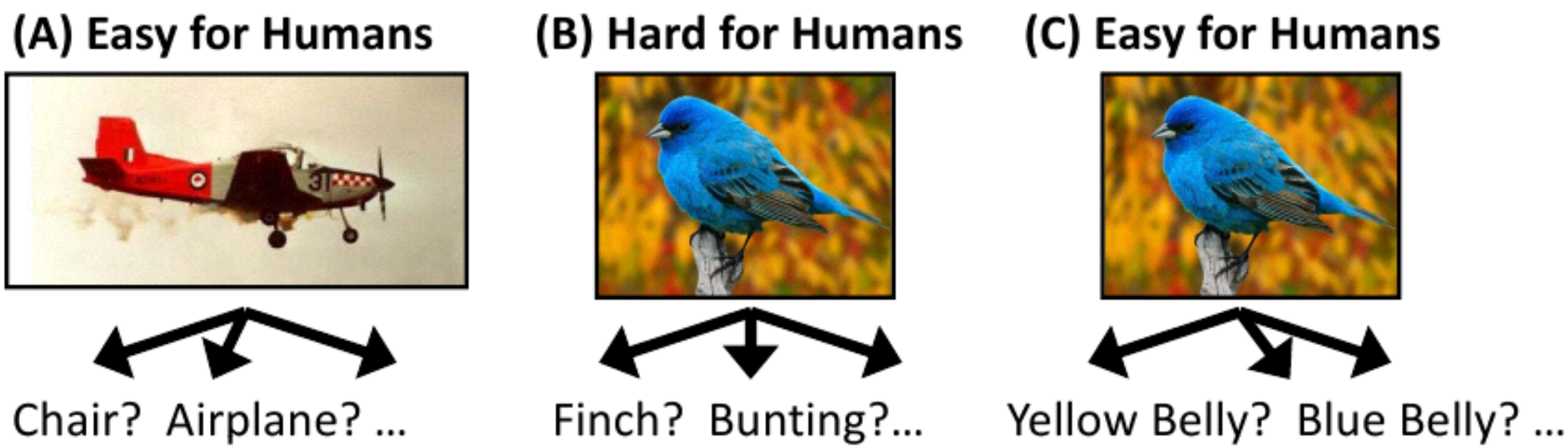
J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, [Real-Time Human Pose Recognition in Parts from a Single Depth Image](), CVPR 2011

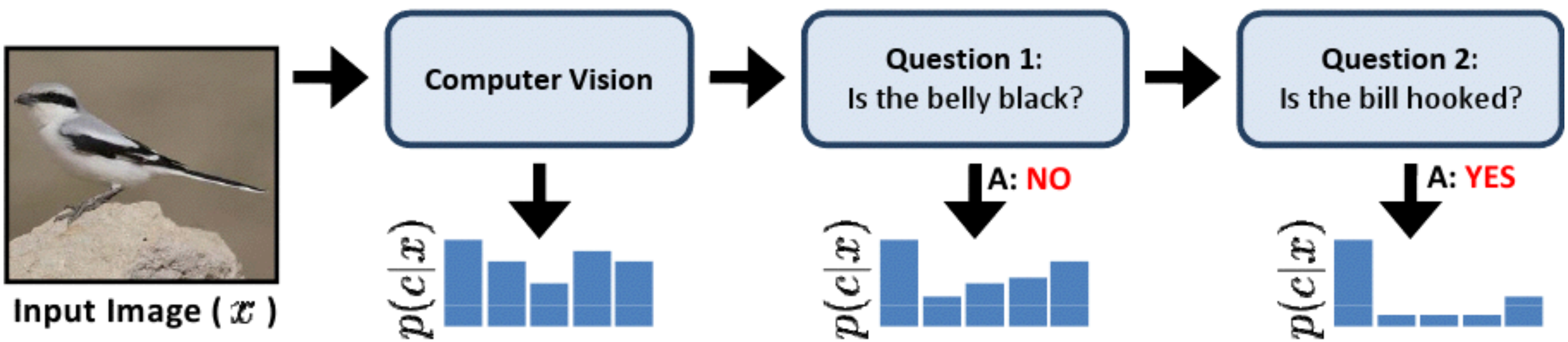
Attributes for recognition



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

Human “in the loop” recognition





Branson S., Wah C., Babenko B., Schroff F., Welinder P., Perona P., Belongie S., “Visual Recognition with Humans in the Loop”, European Conference on Computer Vision (ECCV), Heraklion, Crete, Sept., 2010.

Crowdsourcing

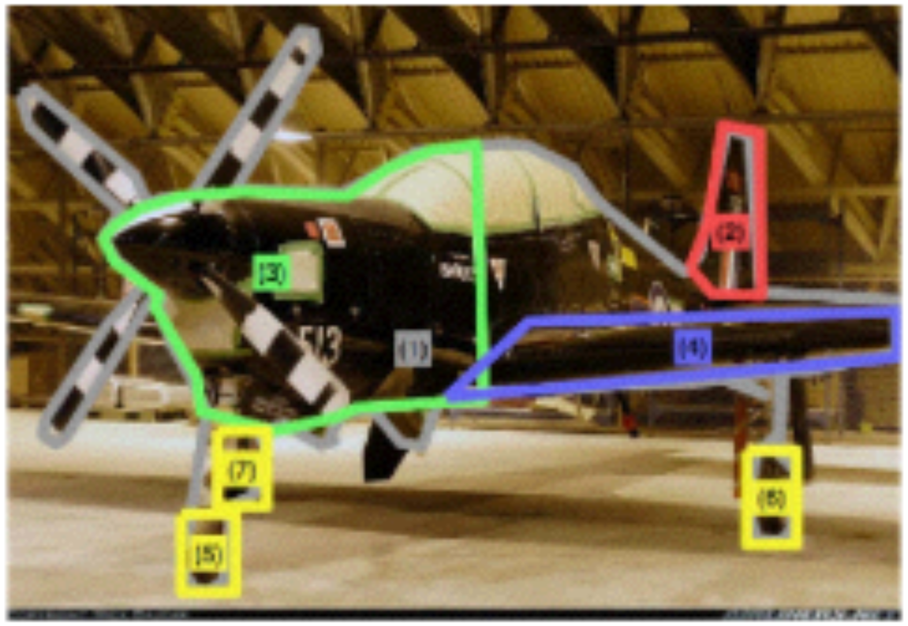
- Large datasets become the norm (real world settings)
  - LabelMe, PASCAL VOC, ImageNet
  - Enable new machine learning methods (e.g., deep learning)

http://www.blogging4jobs.com/hr/solve-your-workplace-issues-by-crowdsourcing/

amazon mechanical turk

Understating objects in detail

OID:Aircraft Benchmark



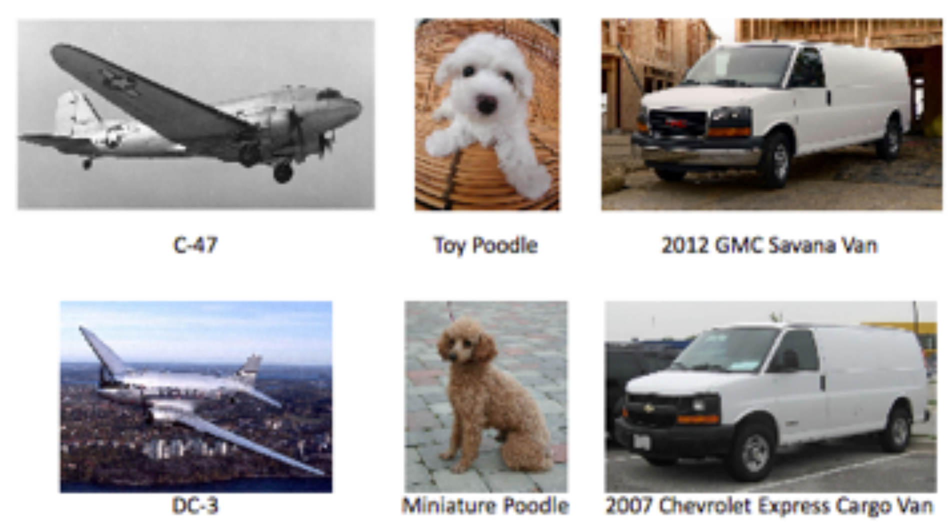
1 **aeroplane** facing-direction: SW; is-airliner: no; is-cargo-plane: no; is-glider: no; is-military-plane: yes; is-propellor-plane: yes; is-seaplane: no; plane-location: on ground/water; plane-size: medium plane; wing-type: single wing plane; undercarriage-arrangement: one-front-two-back; airline: UK–Air Force; model: Short S-312 Tucano T1 2; 2 **vertical stabilizer** tail-has-engine: no-engine 3 **nose** has-engine-or-sensor: has-engine 4 **wing** wing-has-engine: no-engine 5 **undercarriage** cover-type: retractable; group-type: 1-wheel-1-axle; location: front-middle 6 **undercarriage** cover-type: retractable; group-type: 1-wheel-1-axle; location: back-left 7 **undercarriage** cover-type: retractable; group-type: 1-wheel-1-axle; location: back-right.

Vedaldi et al., CVPR 14

Fine-grained recognition

many related classes

often confused



Sentence generation

This is a photograph of one sky, one road and one bus. The blue sky is above the gray road. The gray road is near the shiny bus. The shiny bus is near the blue sky.

There are two aeroplanes. The first shiny aeroplane is near the second shiny aeroplane.

There are one cow and one sky. The golden cow is by the blue sky.

There are one dining table, one chair and two windows. The wooden dining table is by the wooden chair, and against the first window, and against the second white window. The wooden chair is by the first window, and by the second white window. The first window is by the second white window.

This is a picture of one sky, one road and one sheep. The gray sky is over the gray road. The gray sheep is by the gray road.

Here we see one road, one sky and one bicycle. The road is near the blue sky, and near the colorful bicycle. The colorful bicycle is within the blue sky.

Here we see two persons, one sky and one aeroplane. The first black person is by the blue sky. The blue sky is near the shiny aeroplane. The second black person is by the blue sky. The shiny aeroplane is by the first black person, and by the second black person.

This is a picture of two dogs. The first dog is near the second furry dog.

G. Kulkarni, V. Premraj, S. Dhar, S. Li, Y. Choi, A. Berg, T. Berg, **Baby Talk: Understanding and Generating Simple Image Descriptions**, CVPR 2011

Deep learning

HOME PAGE | TODAY'S PAPER | VIDEO | MOST POPULAR | U.S. Edition

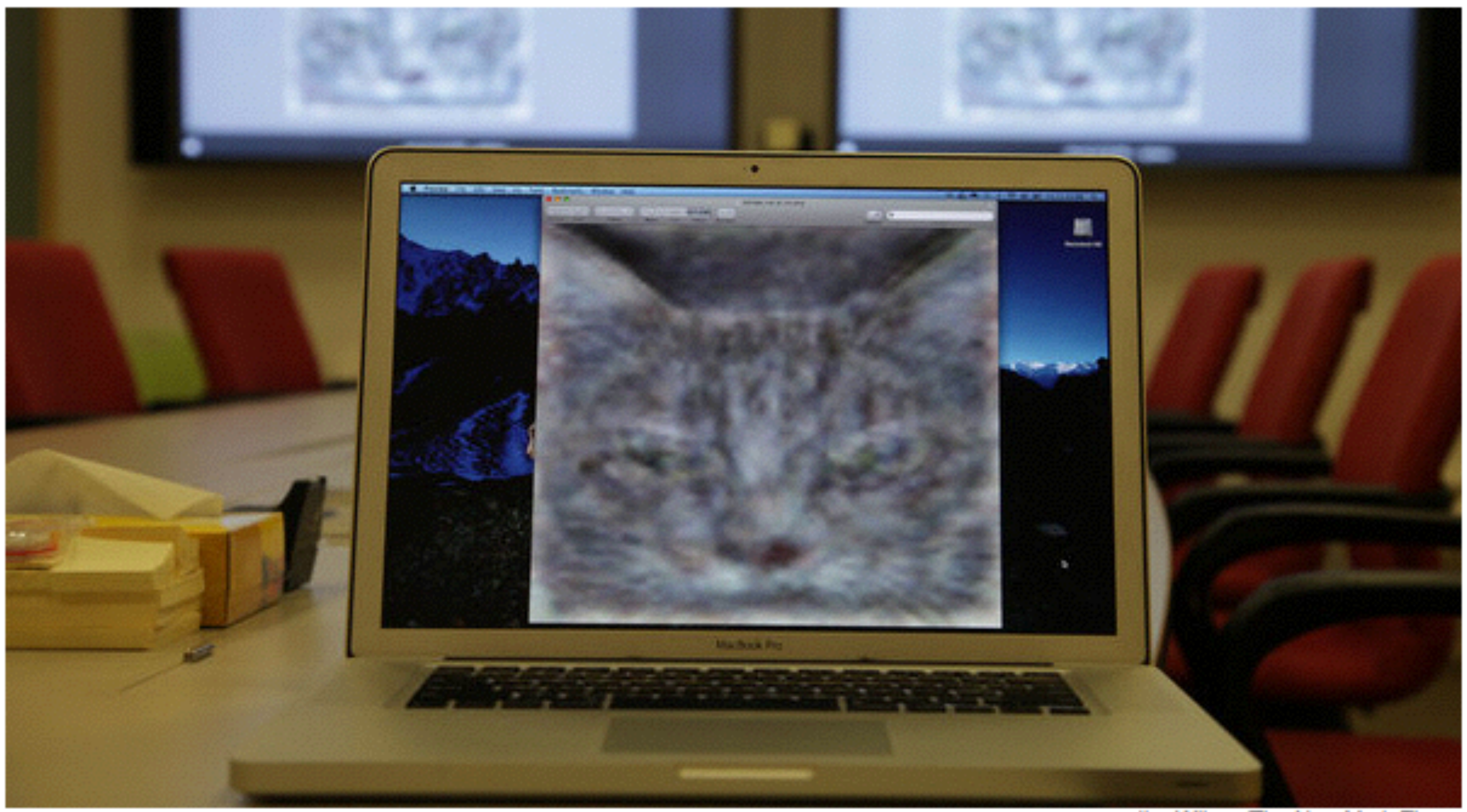
The New York Times

Business Day

Technology

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION

How Many Computers to Identify a Cat? 16,000



Jim Wilson/The New York Times

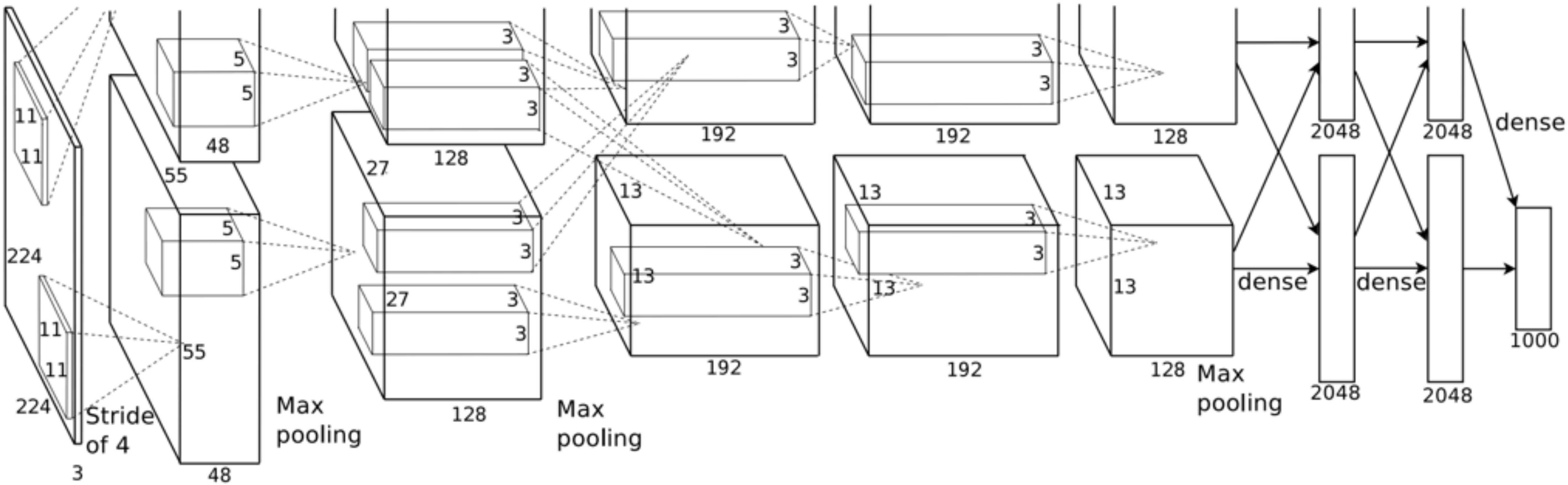
An image of a cat that a neural network taught itself to recognize.

By JOHN MARKOFF

Published: June 25, 2012

[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