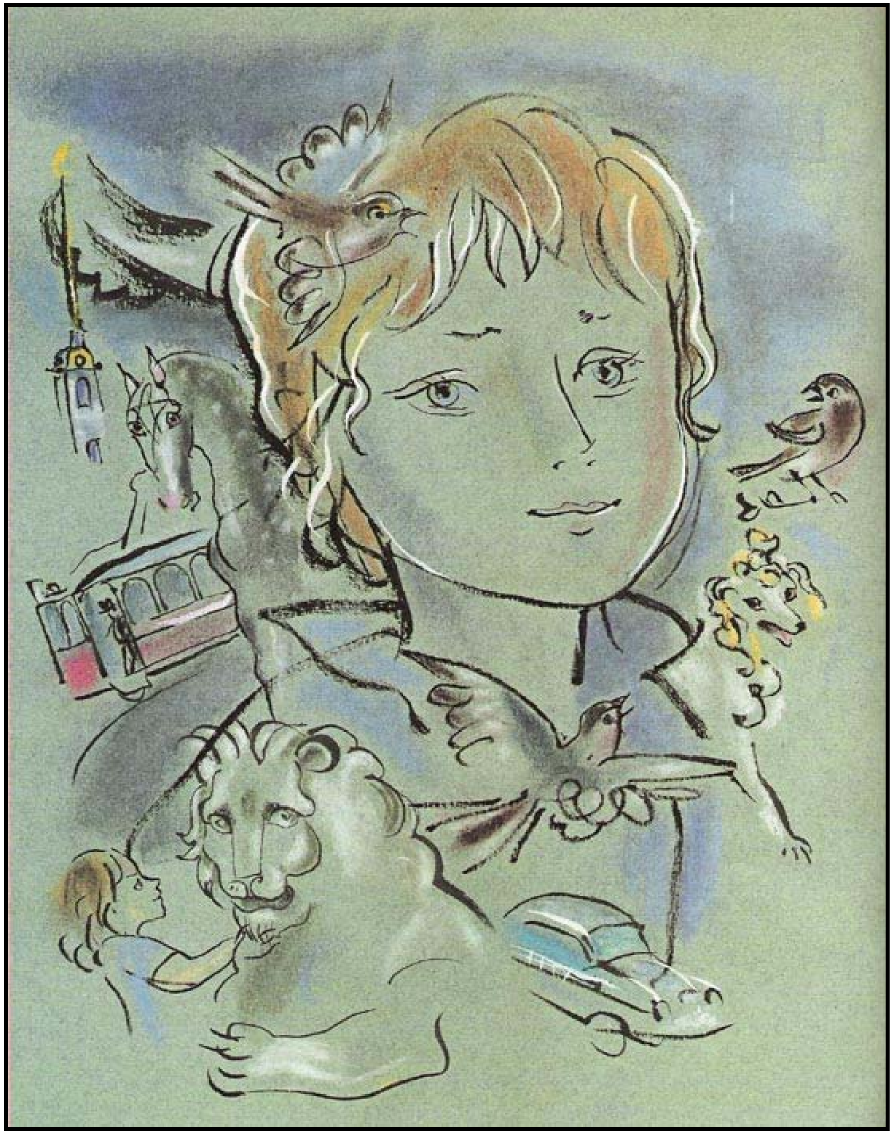
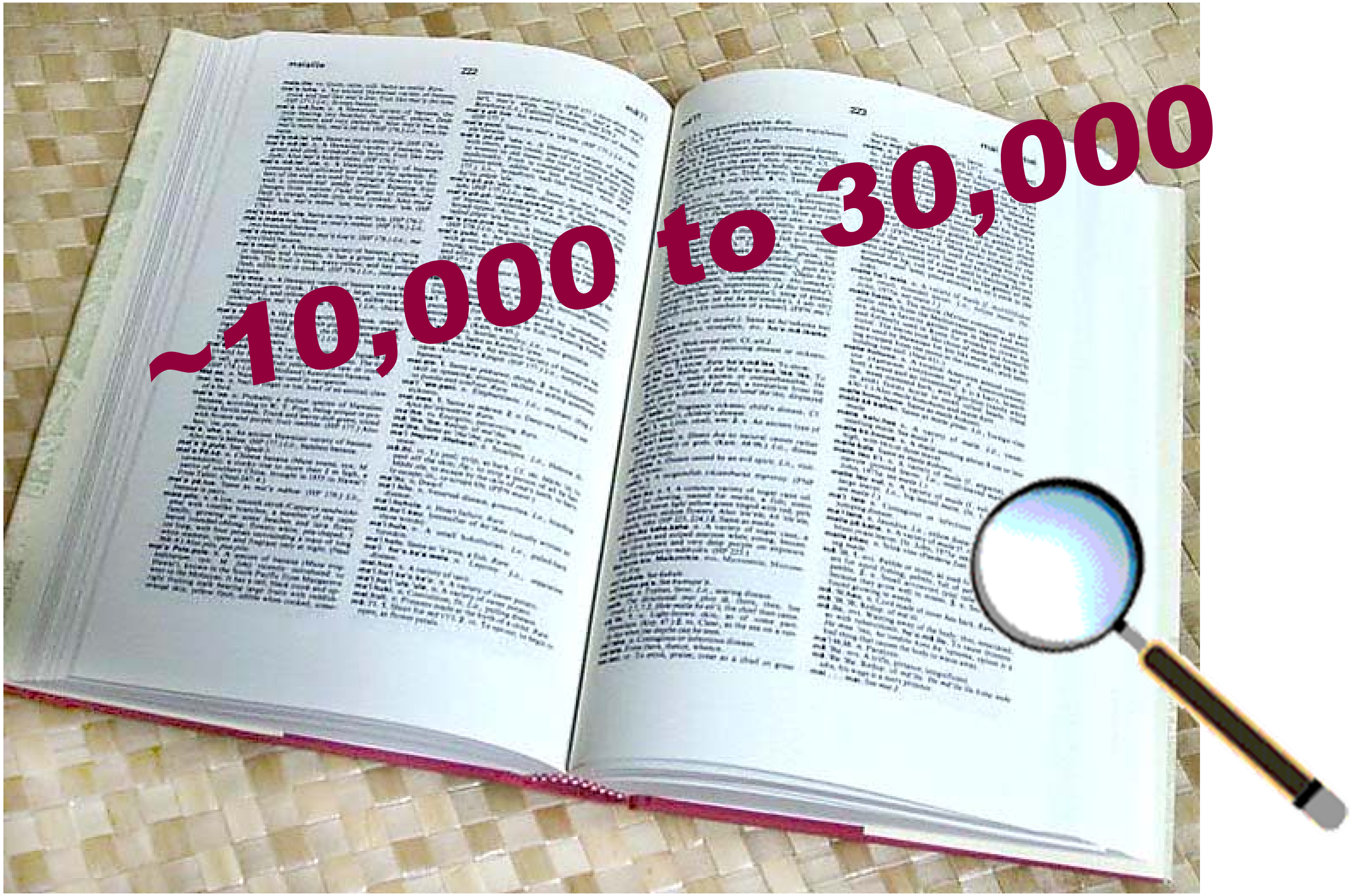
Object Recognition: History and Overview



Slides adapted from Fei-Fei Li, Rob Fergus, Antonio Torralba, and Jean Ponce

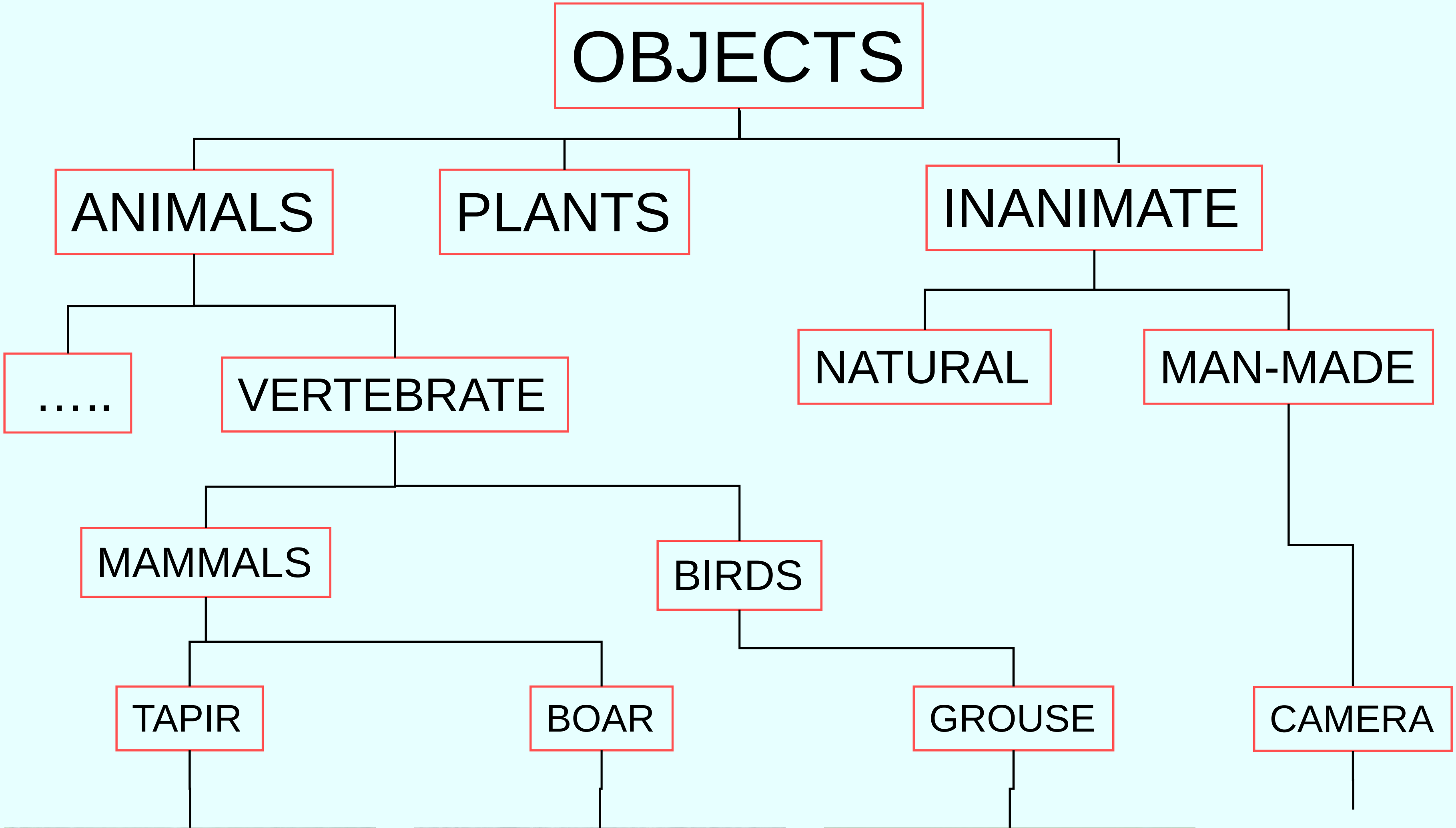
How many visual object categories are there?



~10,000 to 30,000

Biederman 1987

~10,000 to 30,000



OBJECTS

ANIMALS

PLANTS

INANIMATE

.....

VERTEBRATE

NATURAL

MAN-MADE

MAMMALS

BIRDS

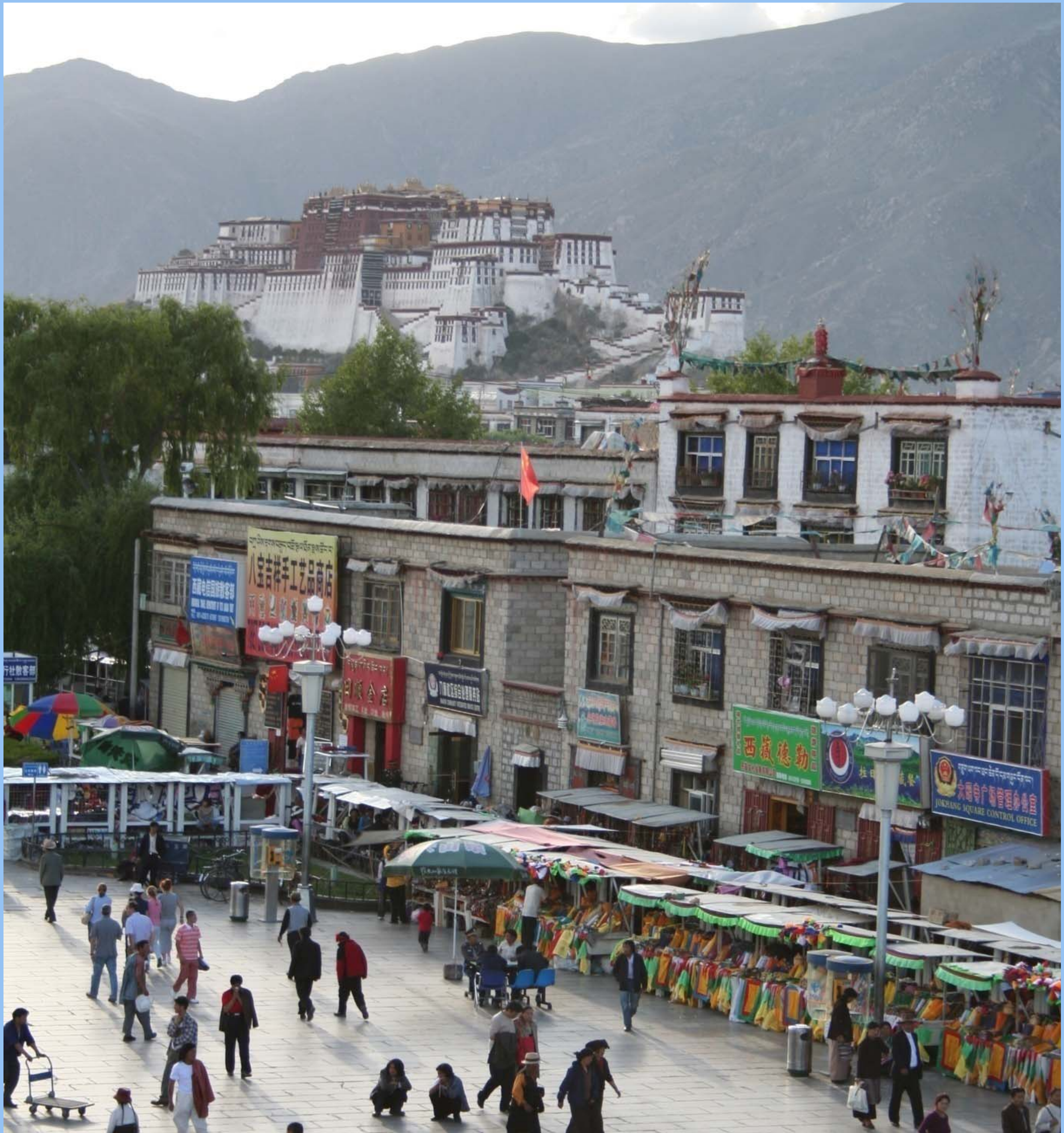
TAPIR

BOAR

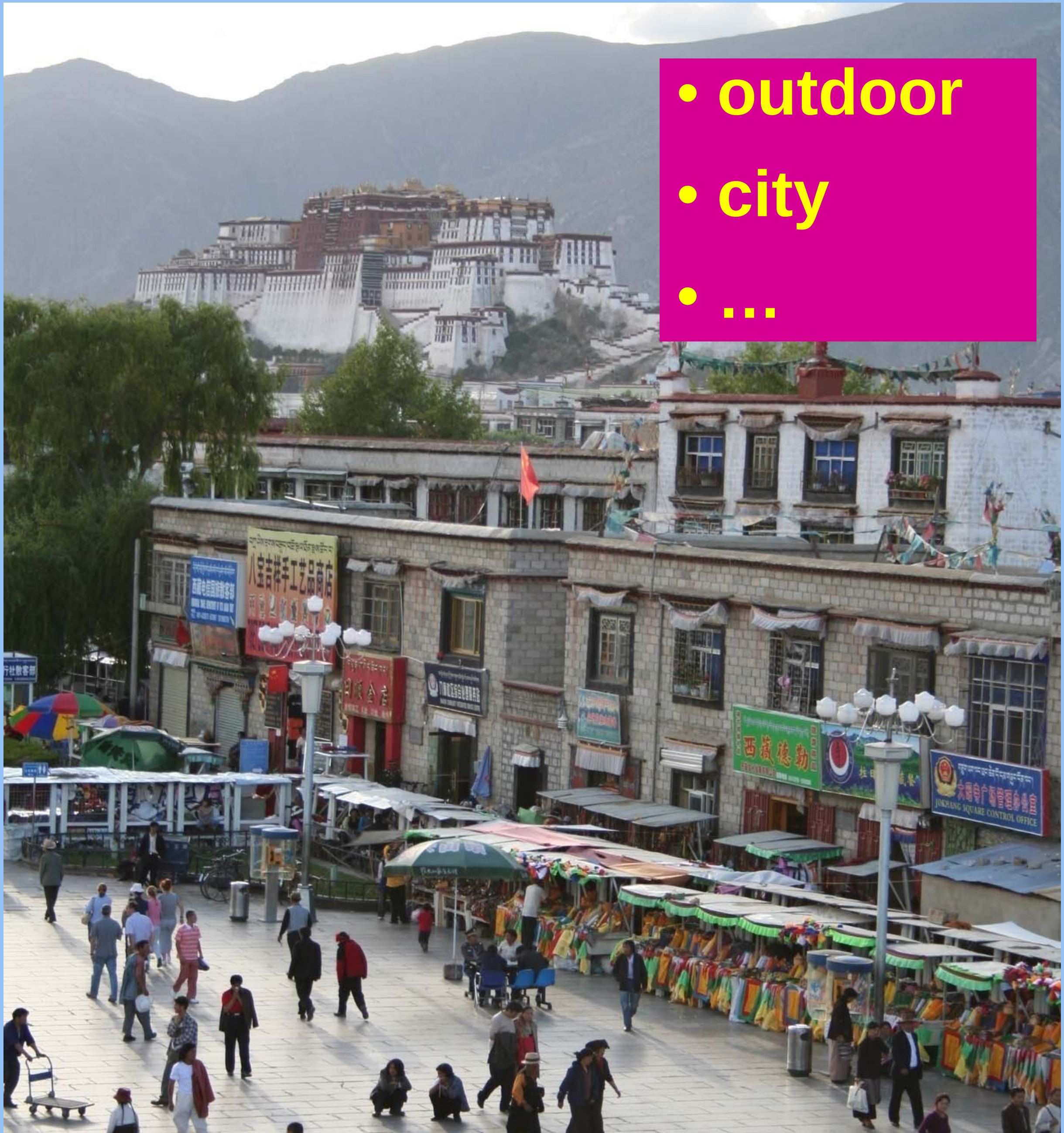
GROUSE

CAMERA

So what does object recognition involve?

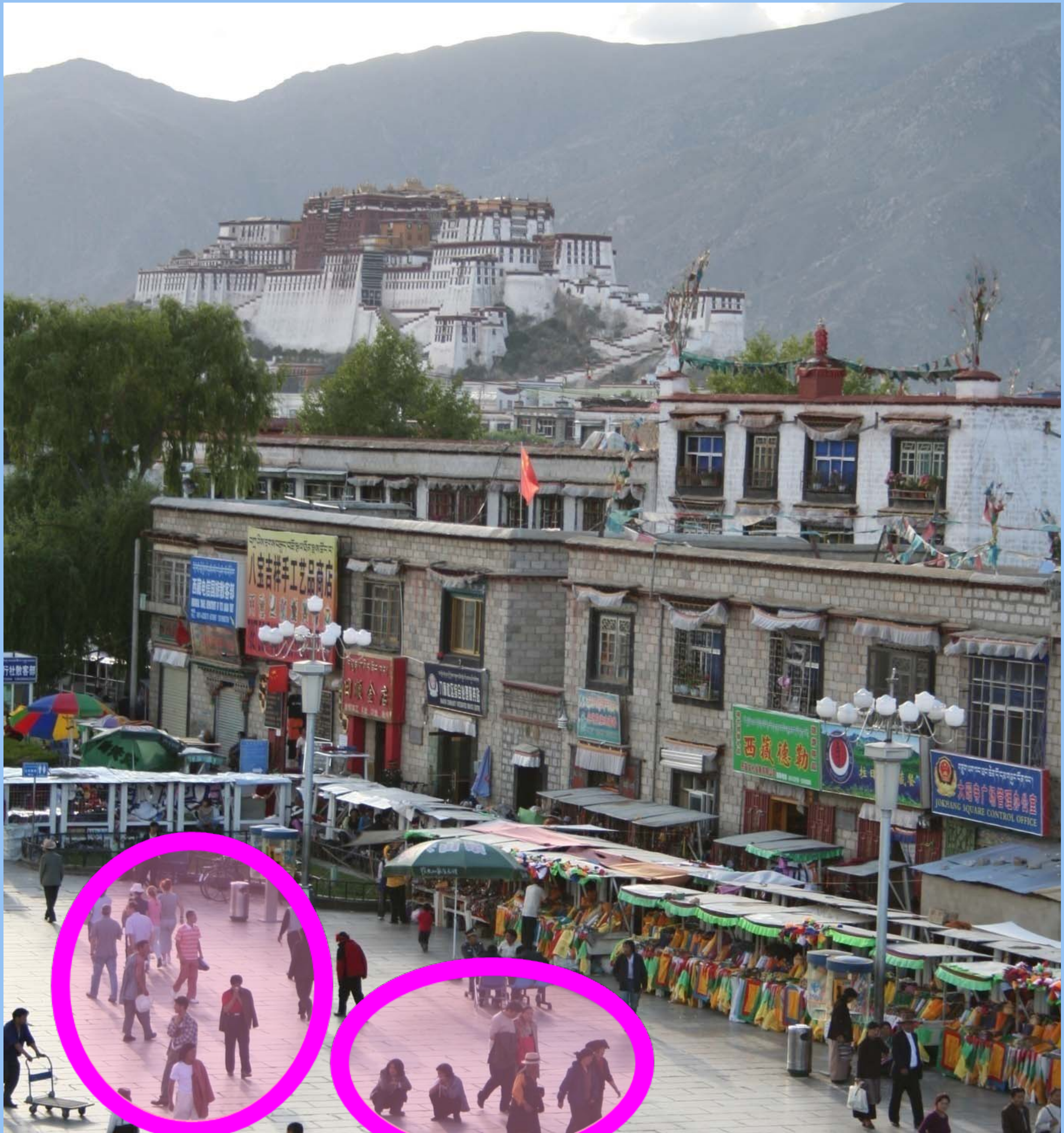


Scene categorization



- **outdoor**
- **city**
- **…**

Object detection: are there people?



Identification: what is this structure?

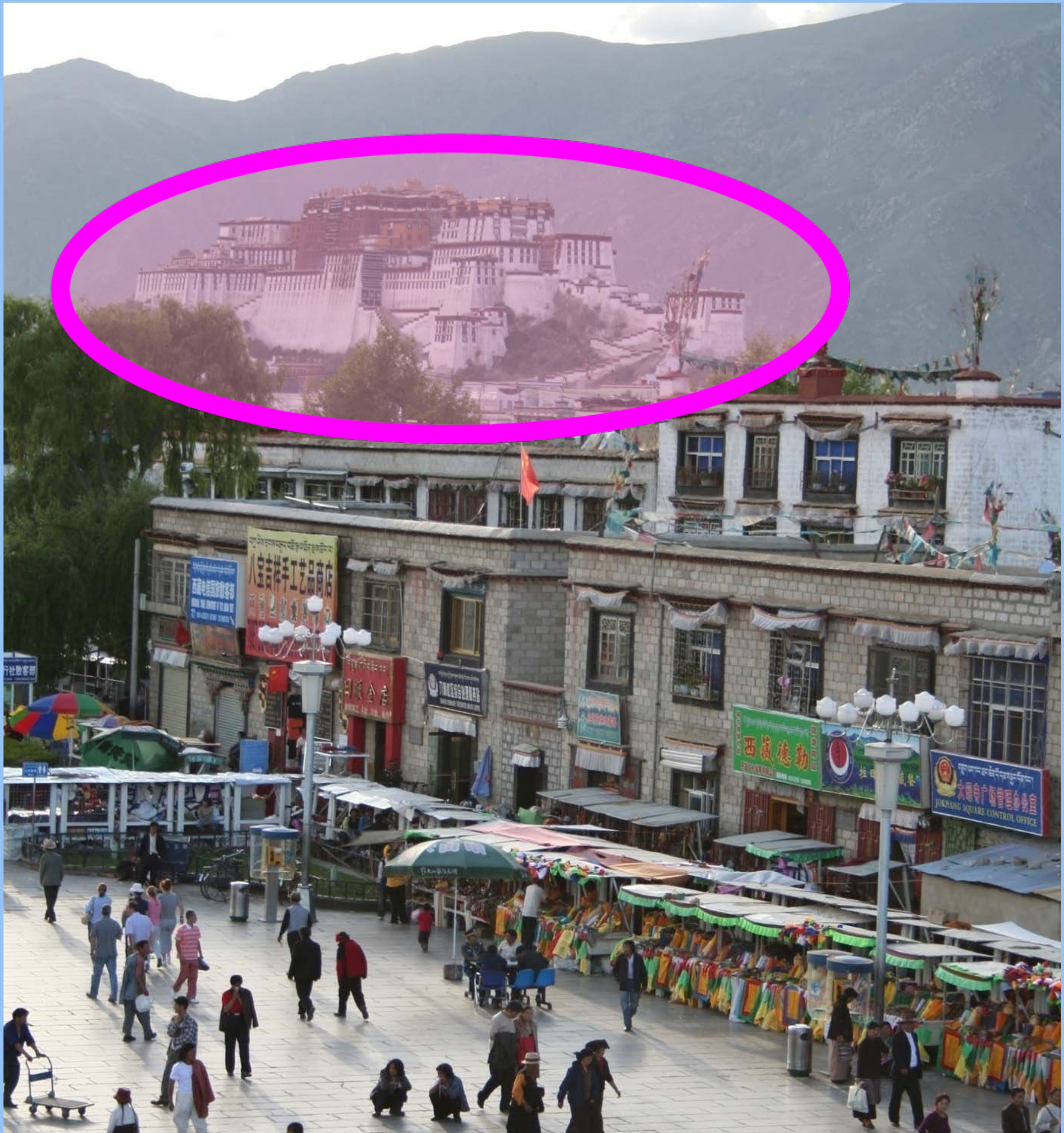
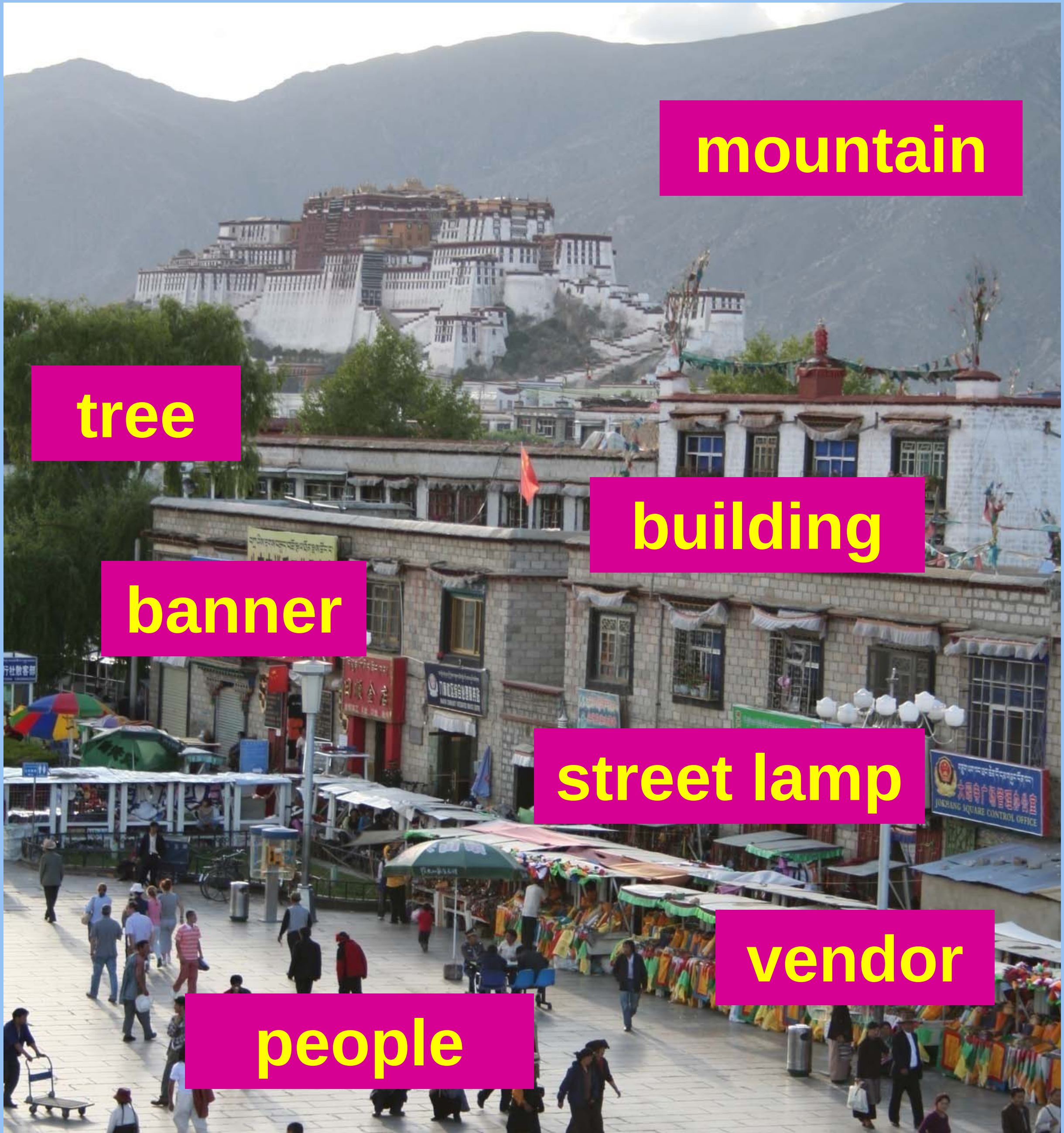
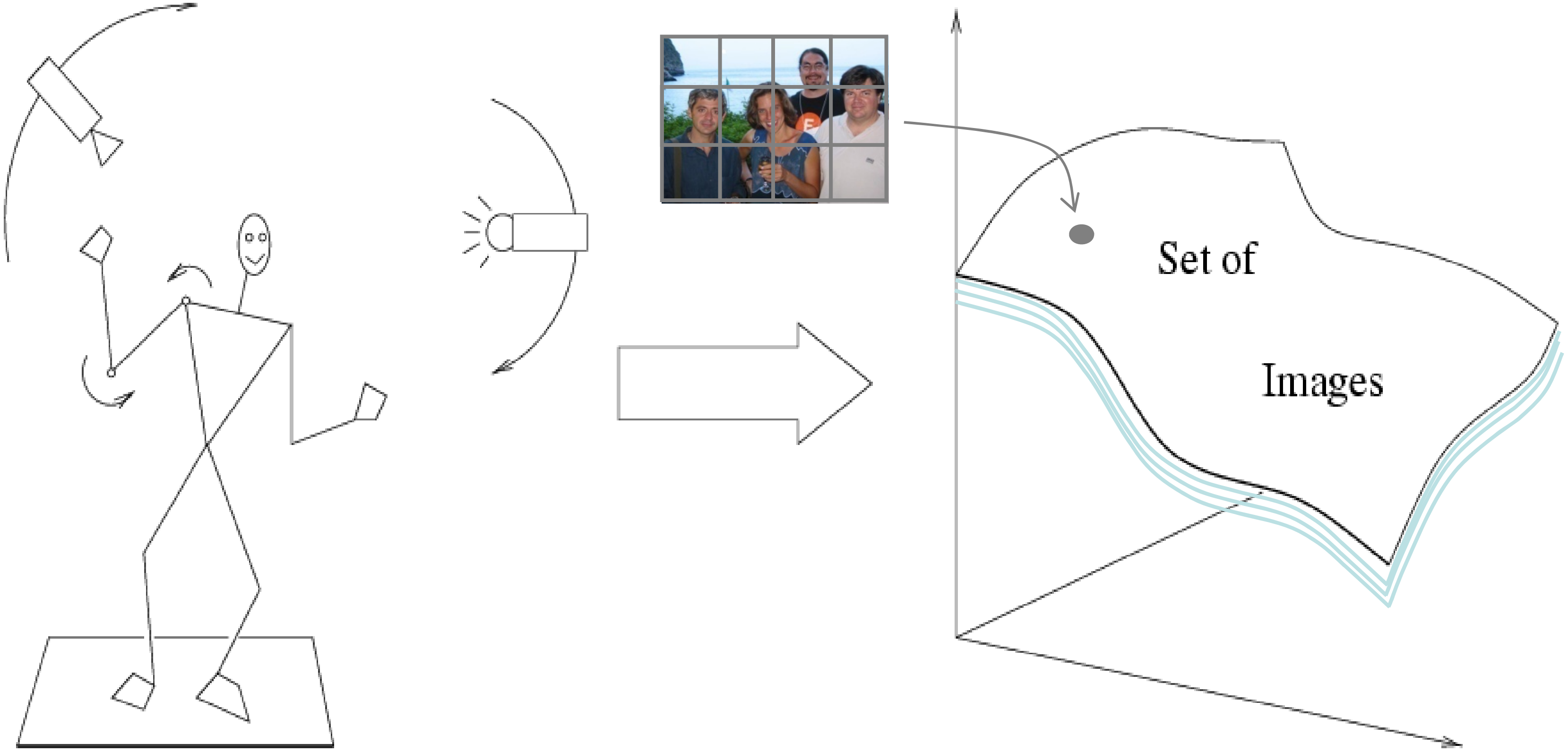


Image parsing



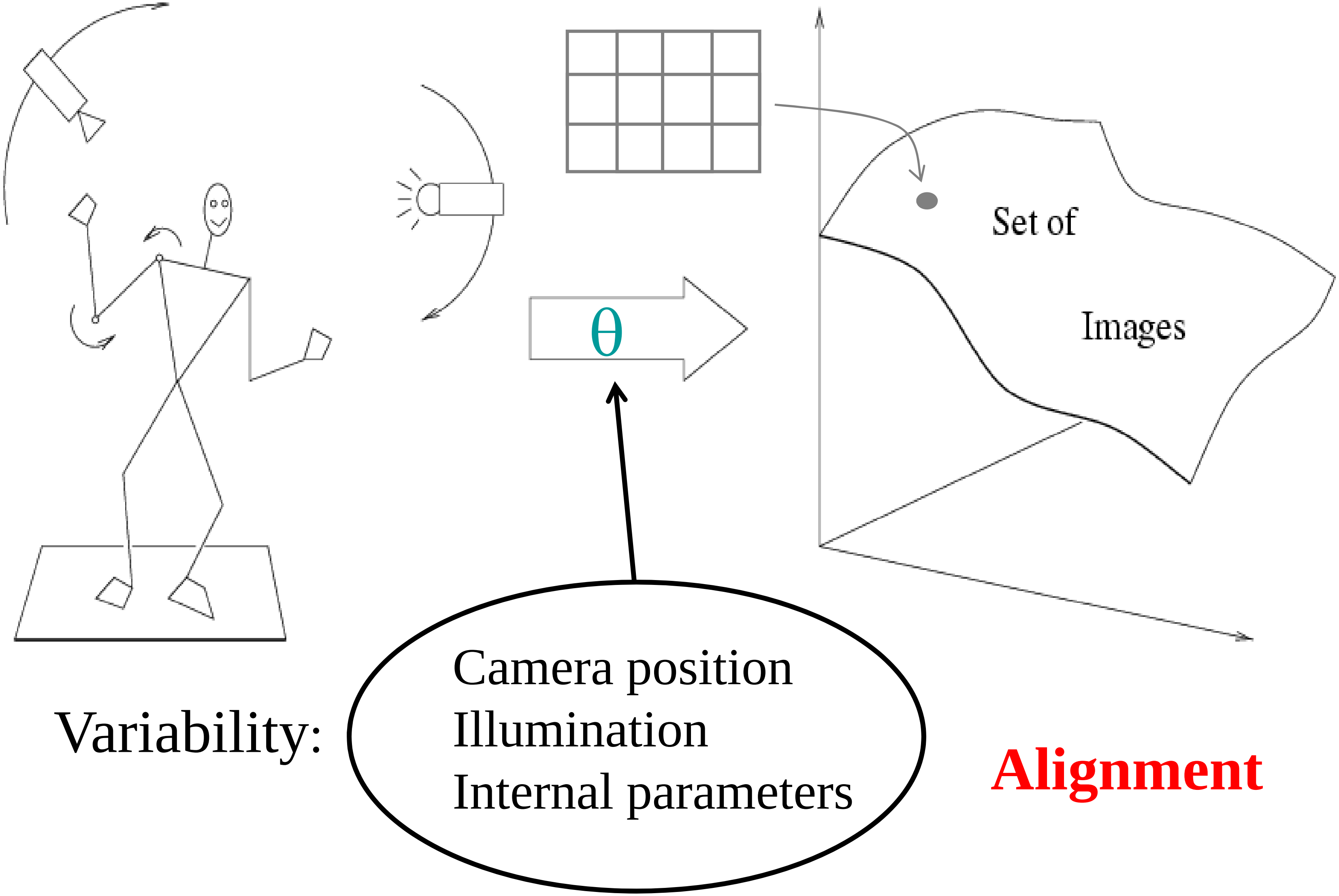
Modeling variability



Variability: Camera position
Illumination
Internal parameters

➡ Within-class variations

Within-class variations



Variability:

Camera position
Illumination
Internal parameters

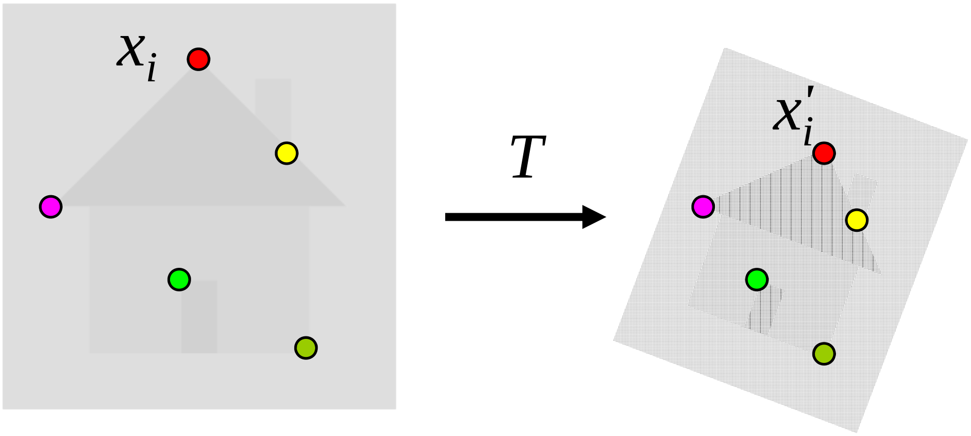
Alignment

Shape: assumed known

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

Recall: Alignment

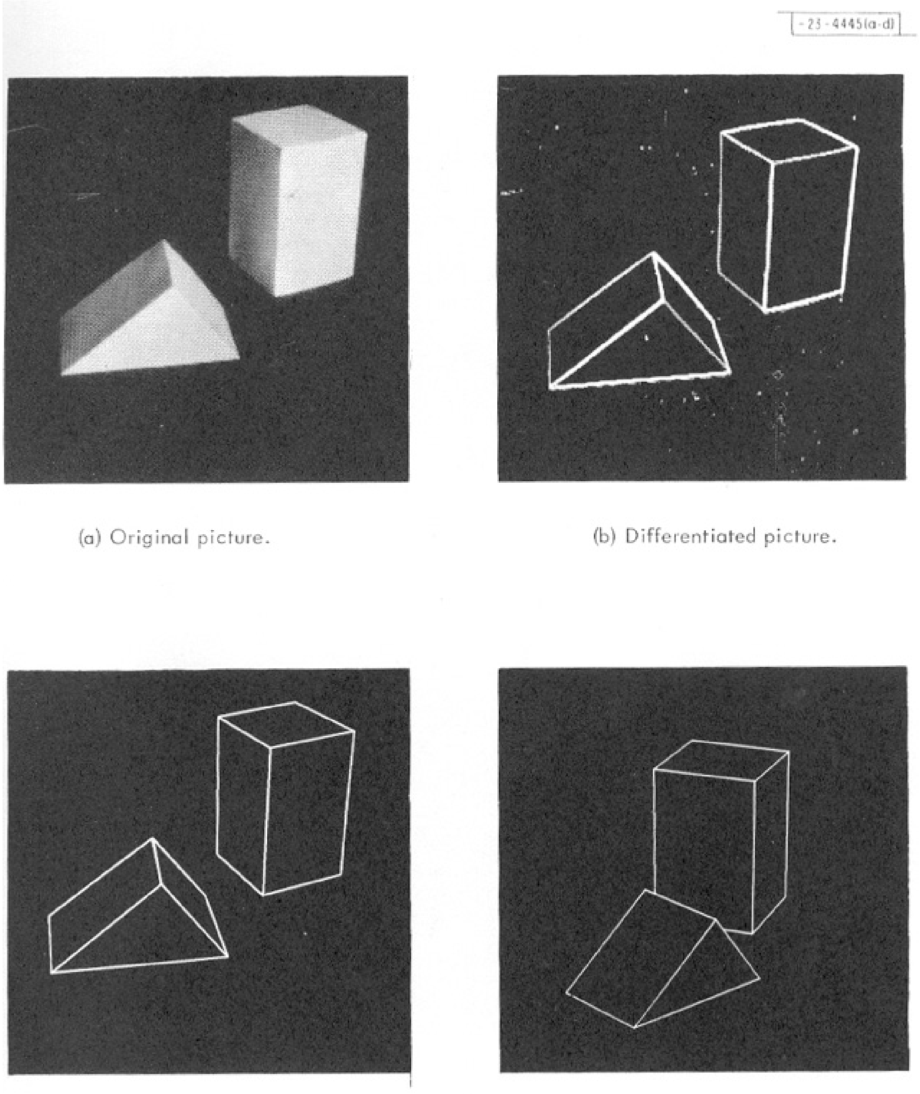
- Alignment: fitting a model to a transformation between pairs of features (*matches*) in two images



Find transformation *T* that minimizes

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

Recall: Origins of computer vision



(a) Original picture.

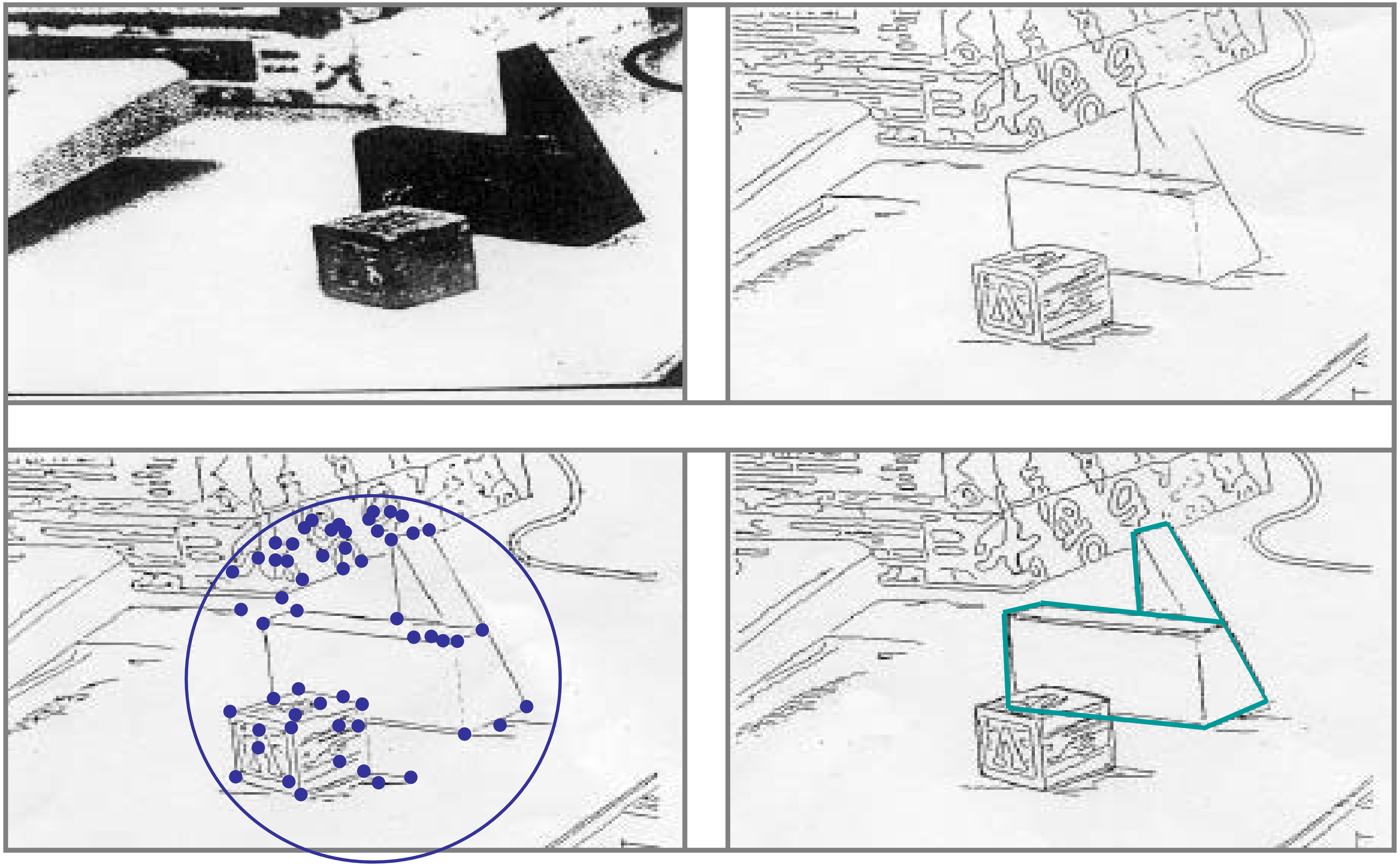
(b) Differentiated picture.

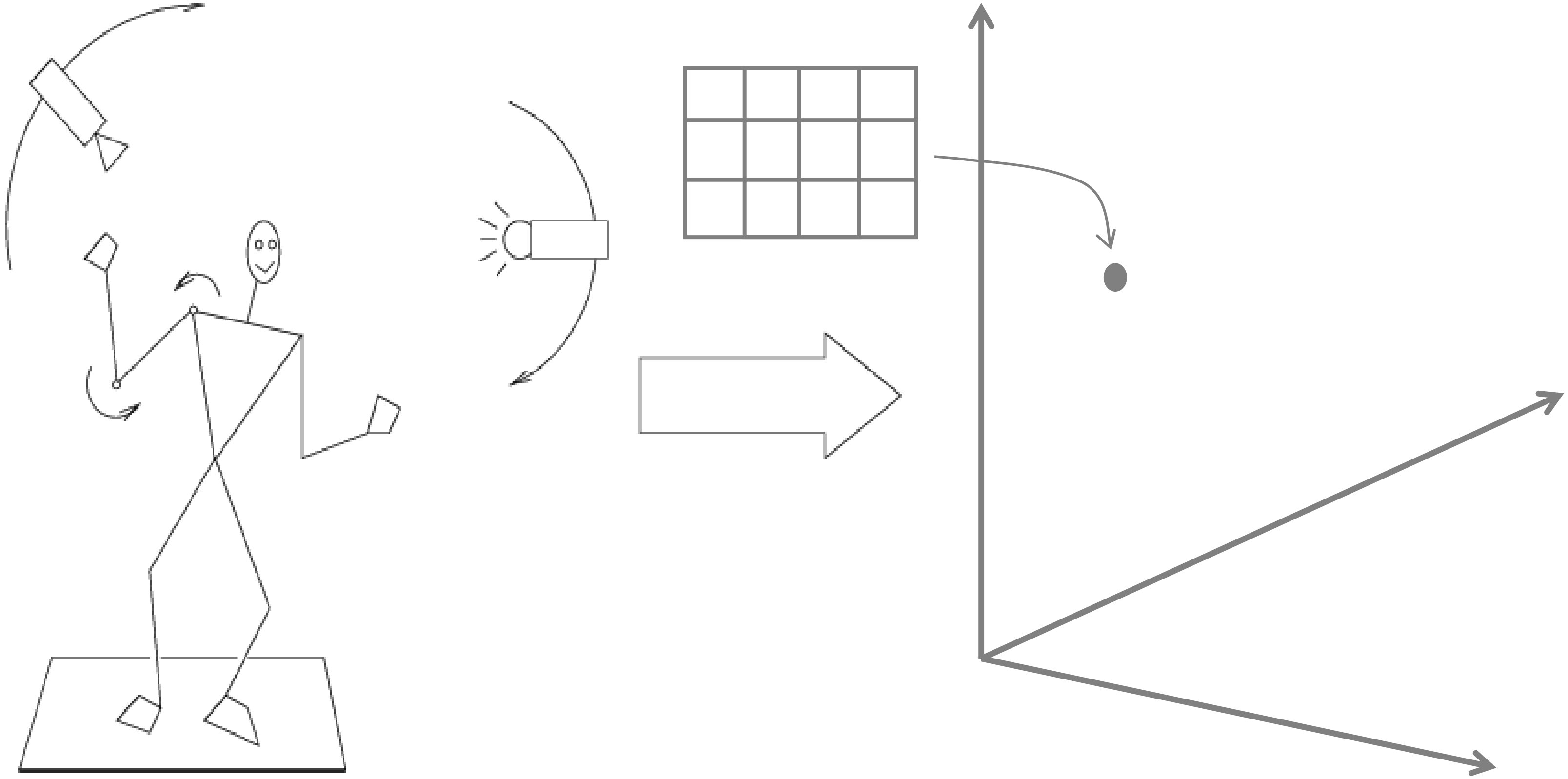
(c) Line drawing.

(d) Rotated view.

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

Alignment: Huttenlocher & Ullman (1987)





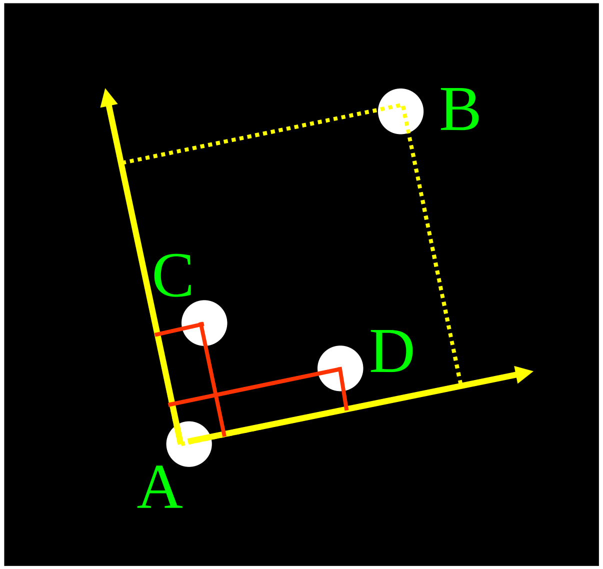
~~Variability~~

Invariance to:

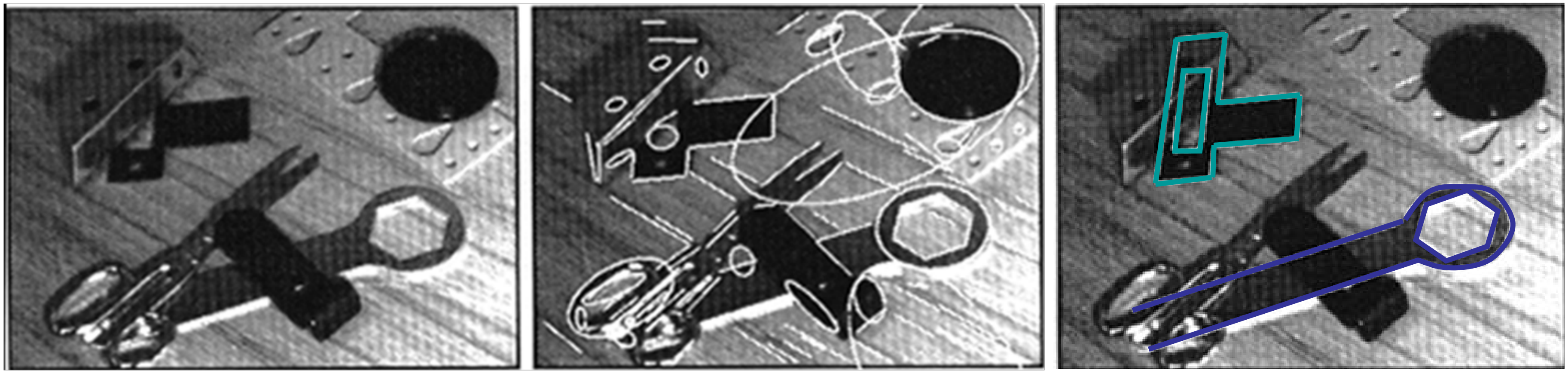
Camera position
Illumination
Internal parameters

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

Example: invariant to similarity transformations computed from four points

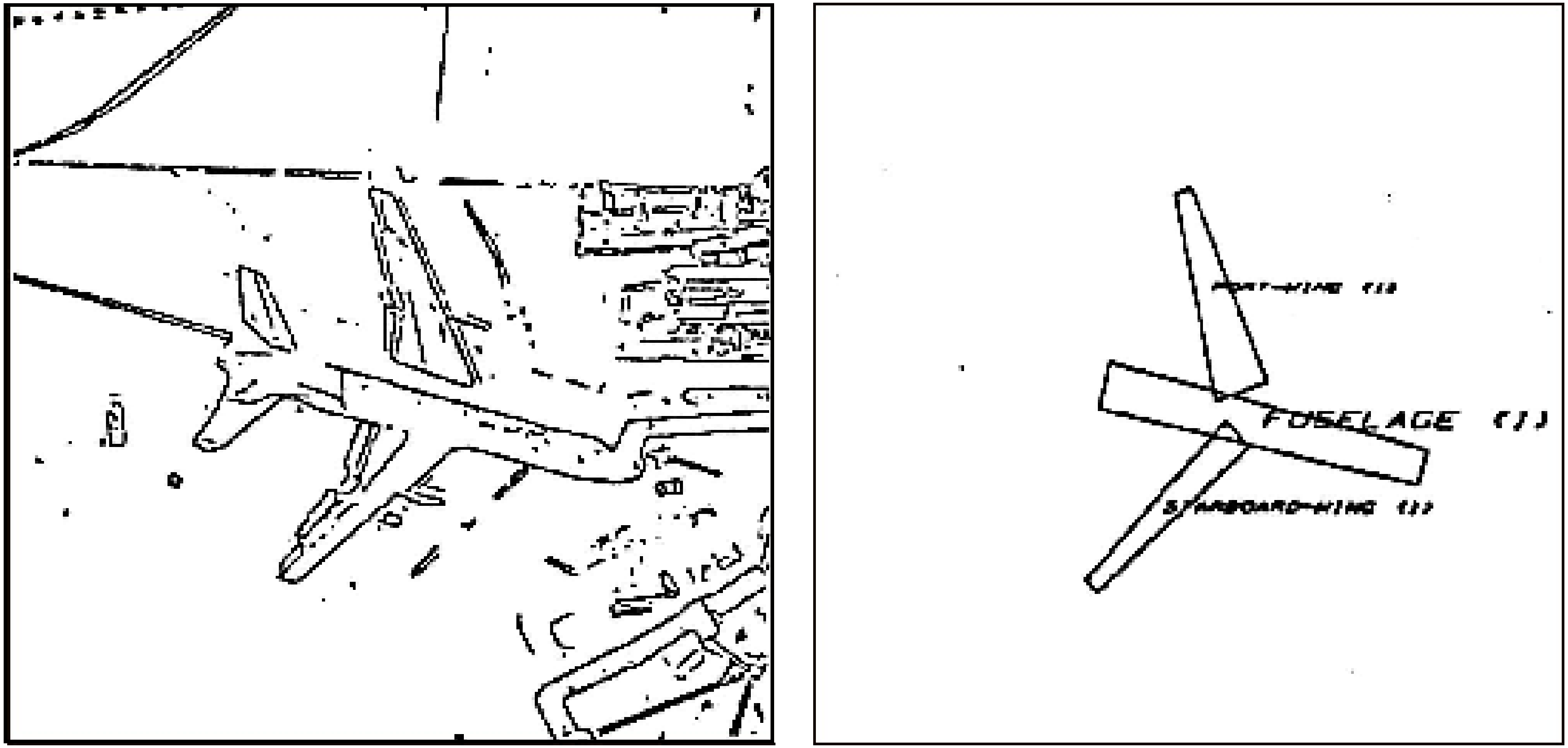


Projective invariants (Rothwell et al., 1992):



General 3D objects do not admit monocular viewpoint invariants (Burns et al., 1993)

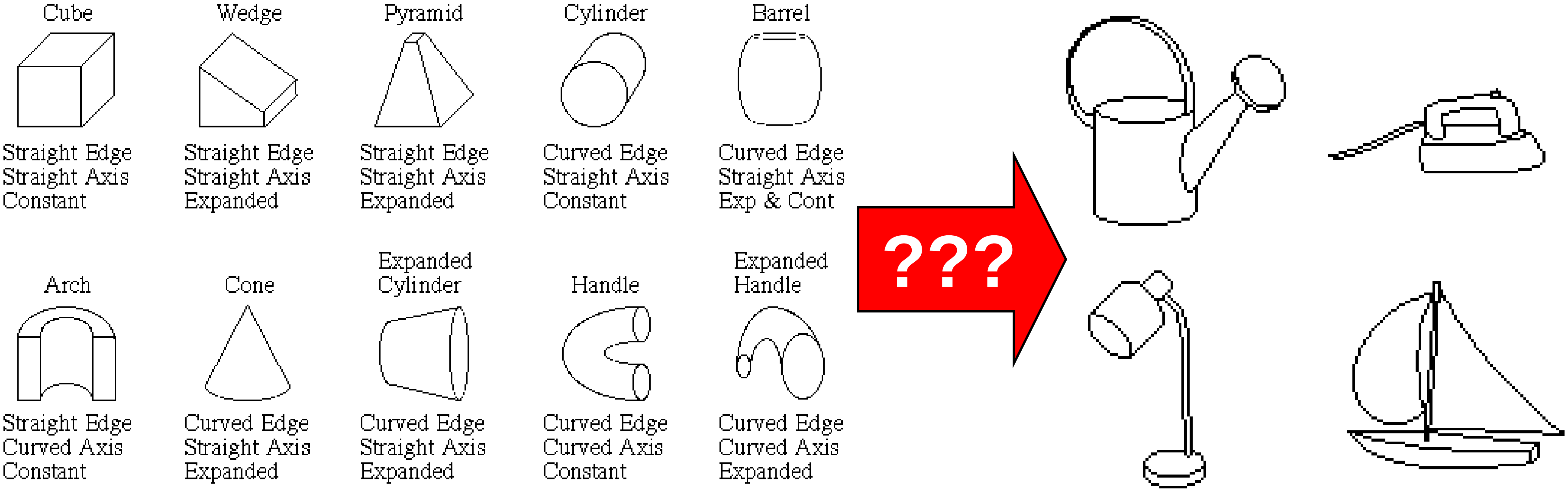
Representing and recognizing object categories is harder...



ACRONYM (Brooks and Binford, 1981)

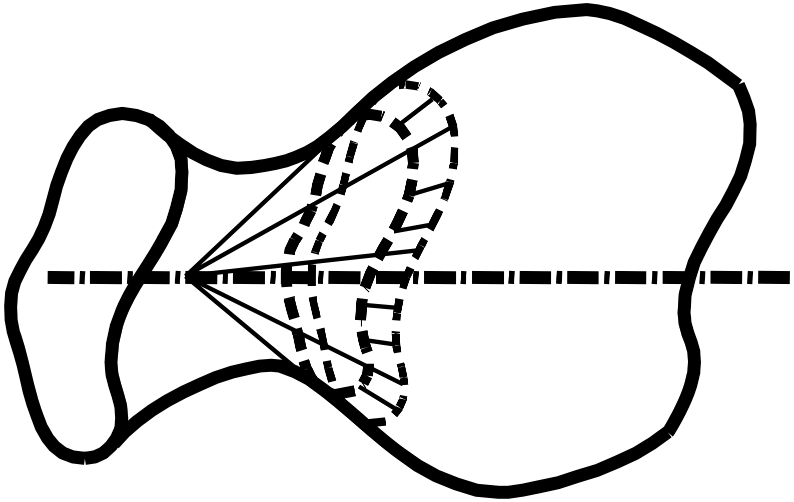
Binford (1971), Nevatia & Binford (1972), Marr & Nishihara (1978)

Recognition by components

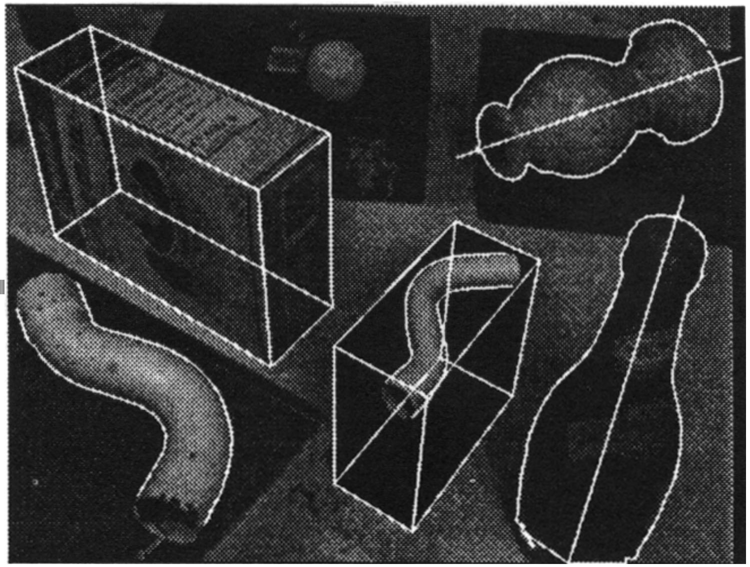


Geons (Biederman 1987)

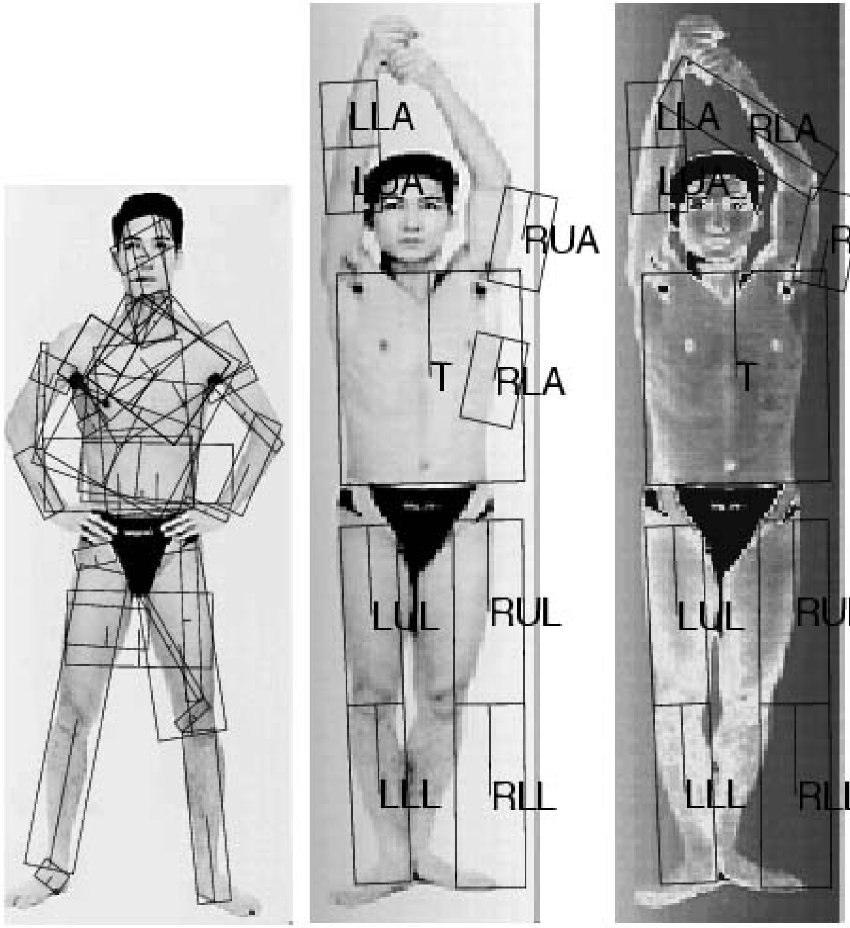
General shape primitives?



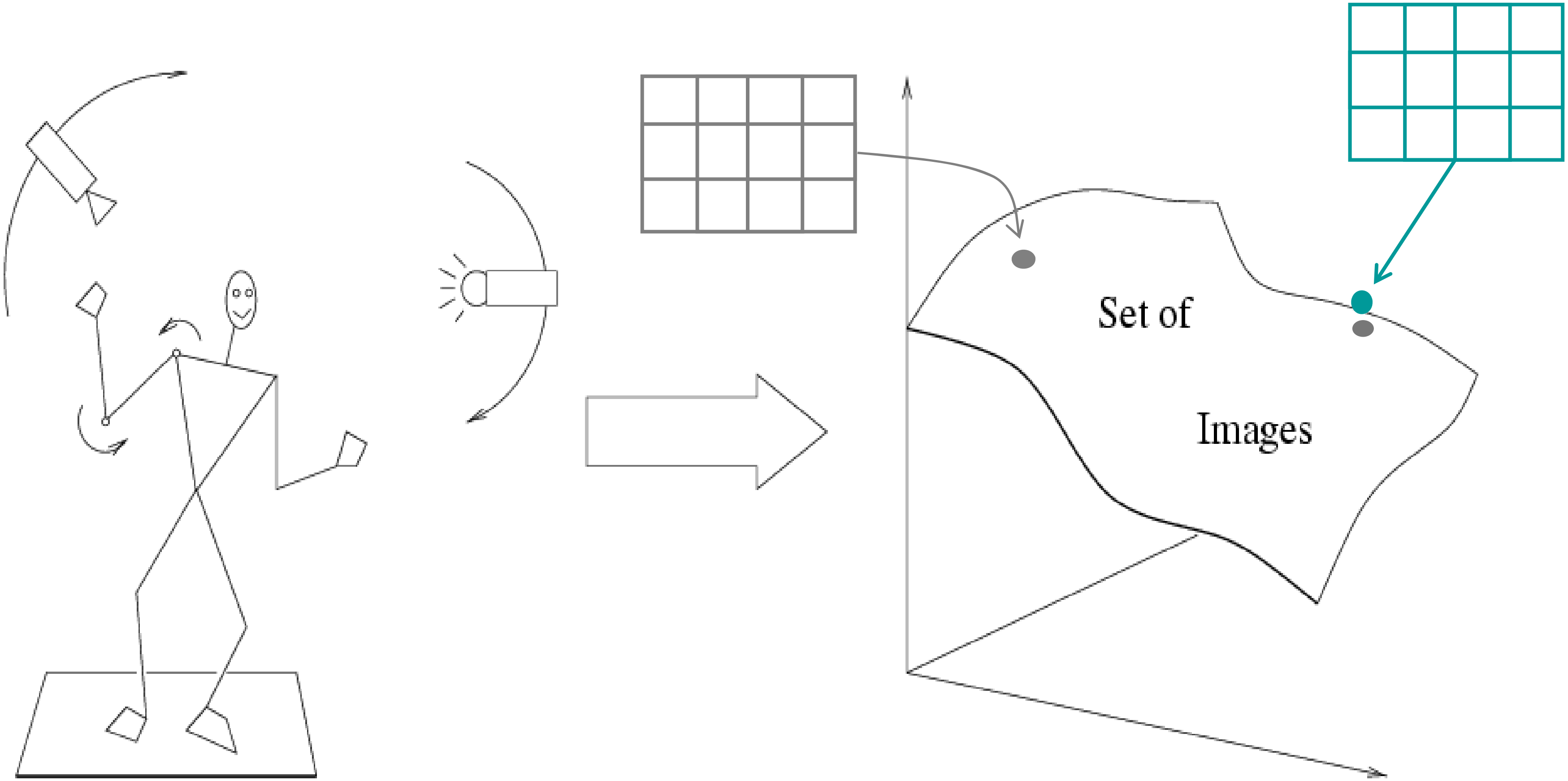
Generalized cylinders
Ponce et al. (1989)



Zisserman et al. (1995)



Forsyth (2000)

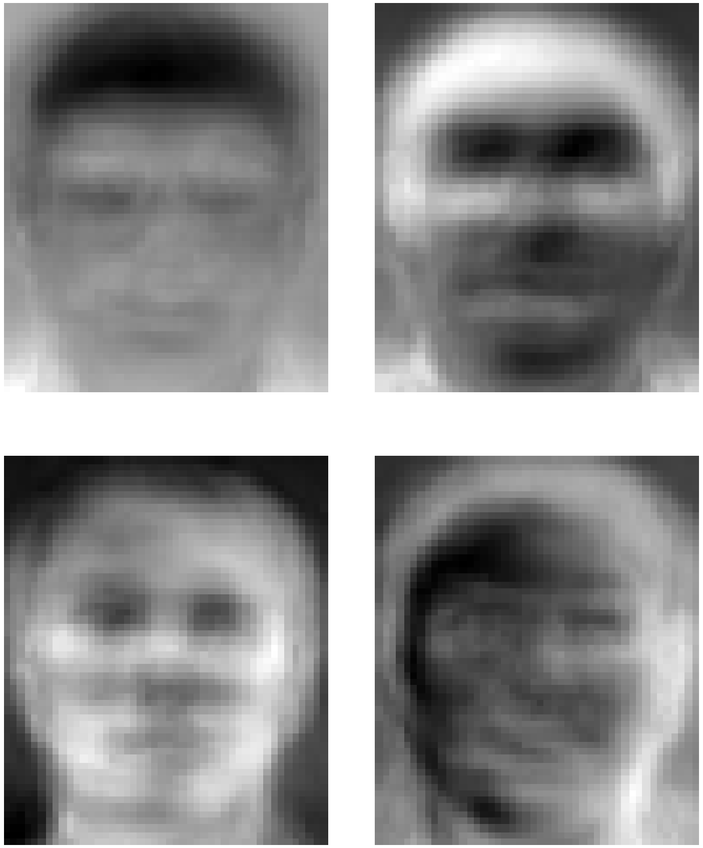


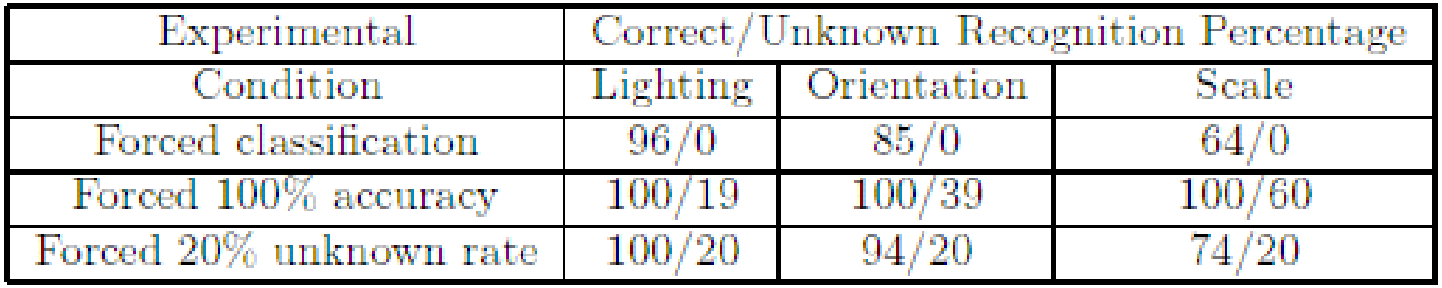
Empirical models of image variability

Appearance-based techniques

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

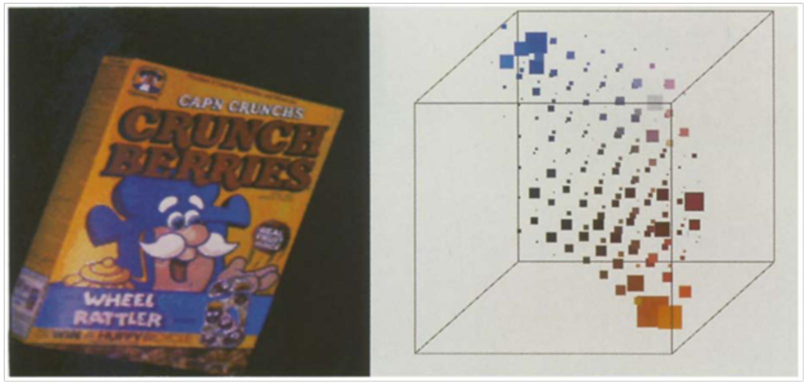
Eigenfaces (Turk & Pentland, 1991)

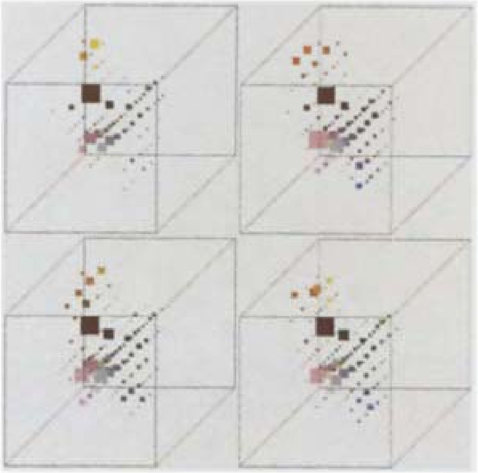




Experimental	Correct/Unknown Recognition Percentage		
Condition	Lighting	Orientation	Scale
Forced classification	96/0	85/0	64/0
Forced 100% accuracy	100/19	100/39	100/60
Forced 20% unknown rate	100/20	94/20	74/20

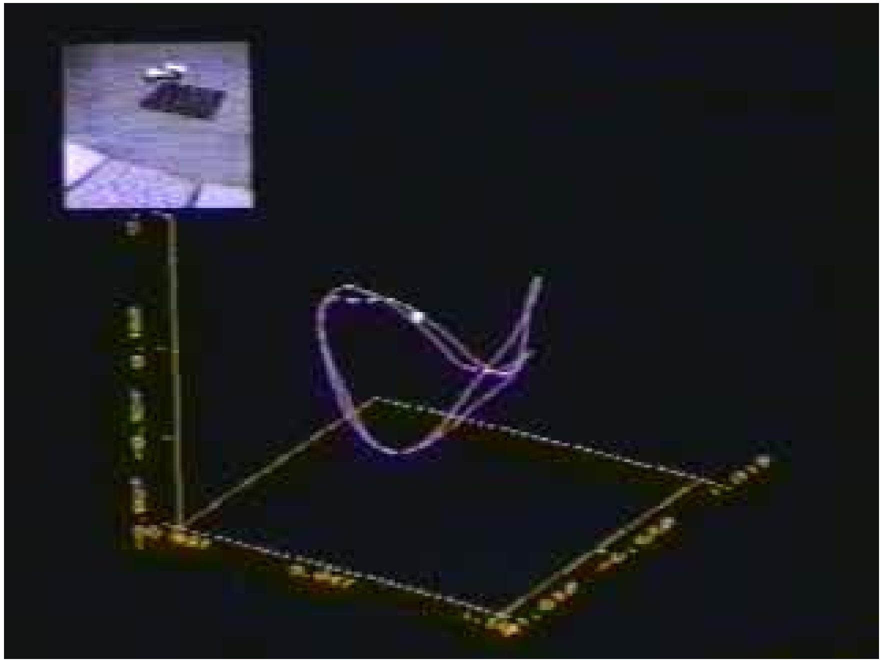
Color Histograms





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

- Can work on relatively simple patterns

- Not robust to clutter, occlusion, lighting changes

Sliding window approaches

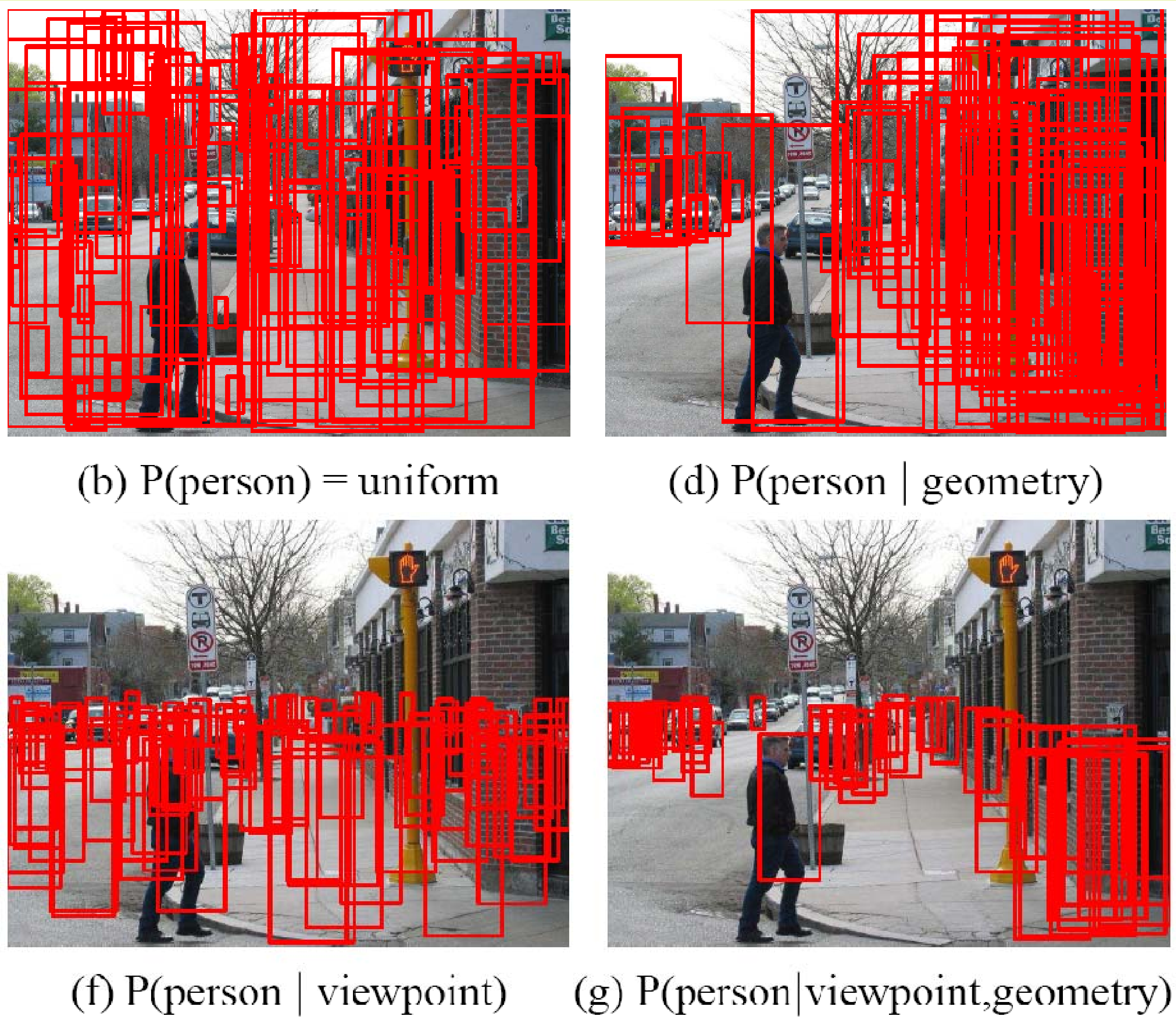
- Turk and Pentland, 1991
- Belhumeur, Hespanha, & Kriegman, 1997
- Schneiderman & Kanade 2004
- Viola and Jones, 2000

- Schneiderman & Kanade, 2004
- Argawal and Roth, 2002
- Poggio et al. 1993

Sliding window approaches

- – Scale / orientation range to search over
- – Speed
- – Context

Context



(b) P(person) = uniform

(d) P(person | geometry)

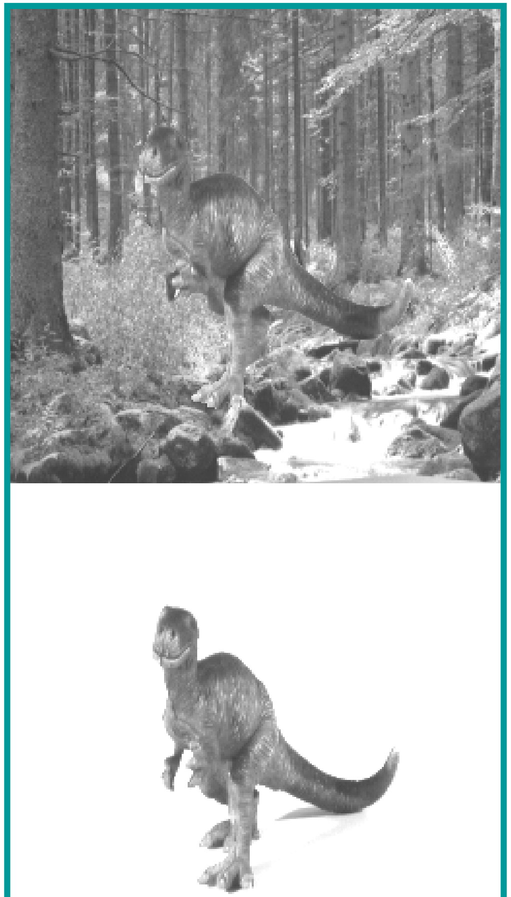
(f) P(person | viewpoint)

(g) P(person|viewpoint,geometry)

Hoiem, Efros, Herbert, 2006

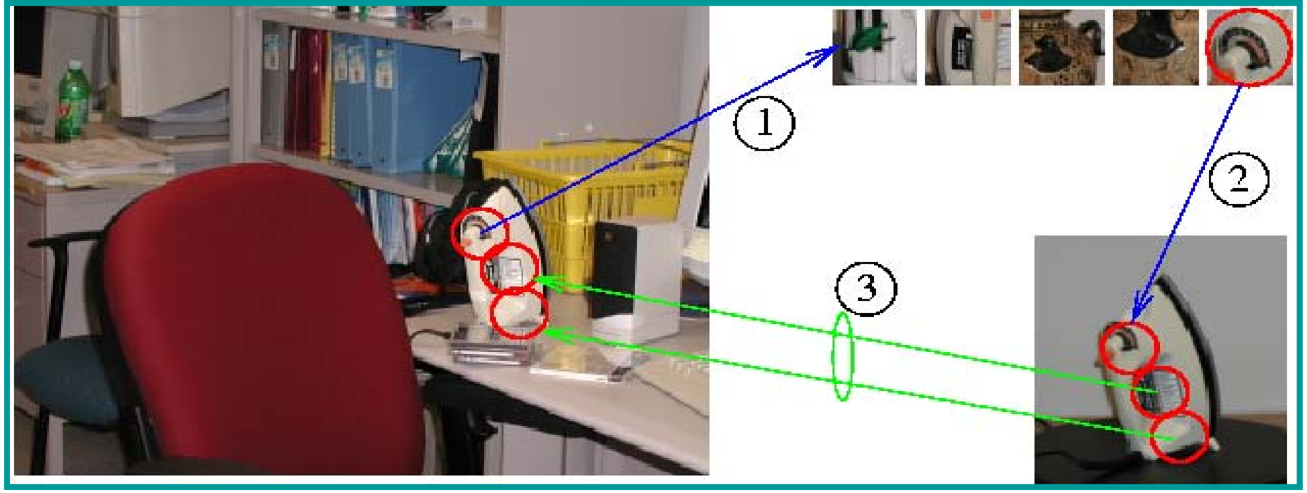
Local features

Combining *local* appearance, spatial constraints, invariants, and classification techniques from machine learning.



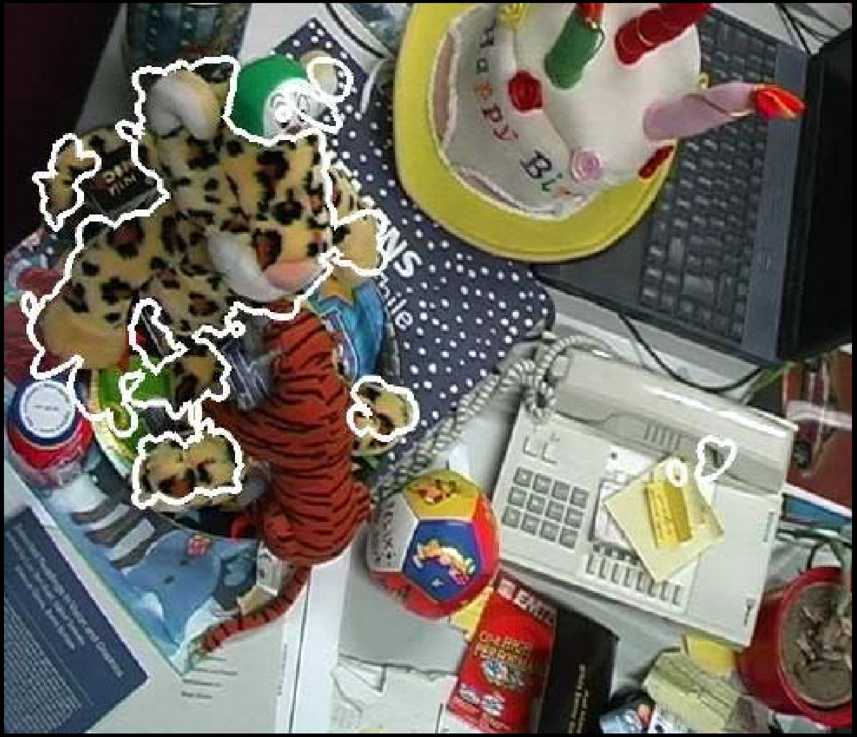
Schmid & Mohr'97

Lowe'02



Mahamud & Hebert'03

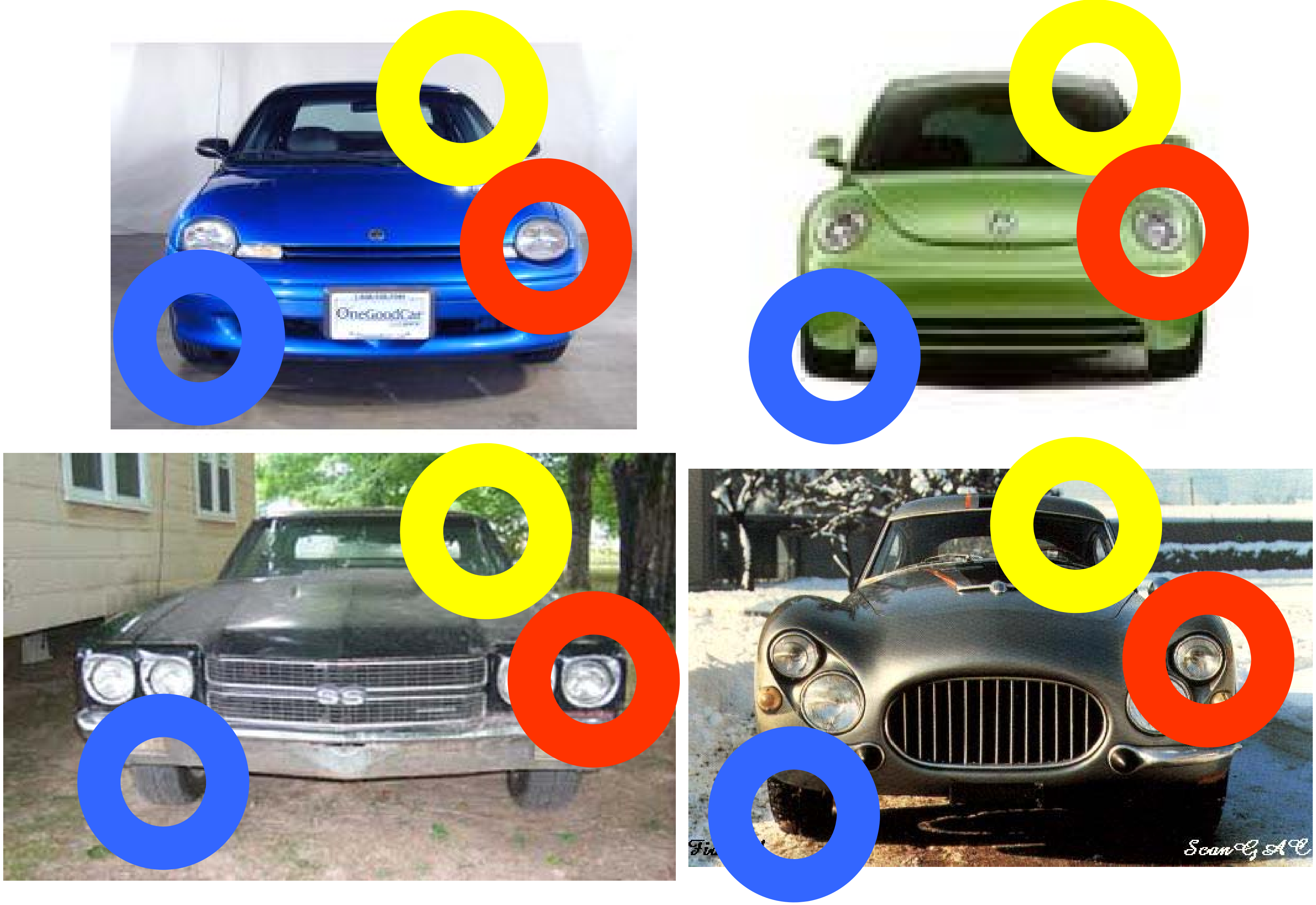
Local features for recognition of object instances



Local features for recognition of object instances

- Lowe, et al. 1999, 2003
- Mahamud and Hebert, 2000
- Ferrari, Tuytelaars, and Van Gool, 2004
- Rothganger, Lazebnik, and Ponce, 2004
- Moreels and Perona, 2005
- …

Representing categories: Parts and Structure



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

Parts-and-shape representation

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

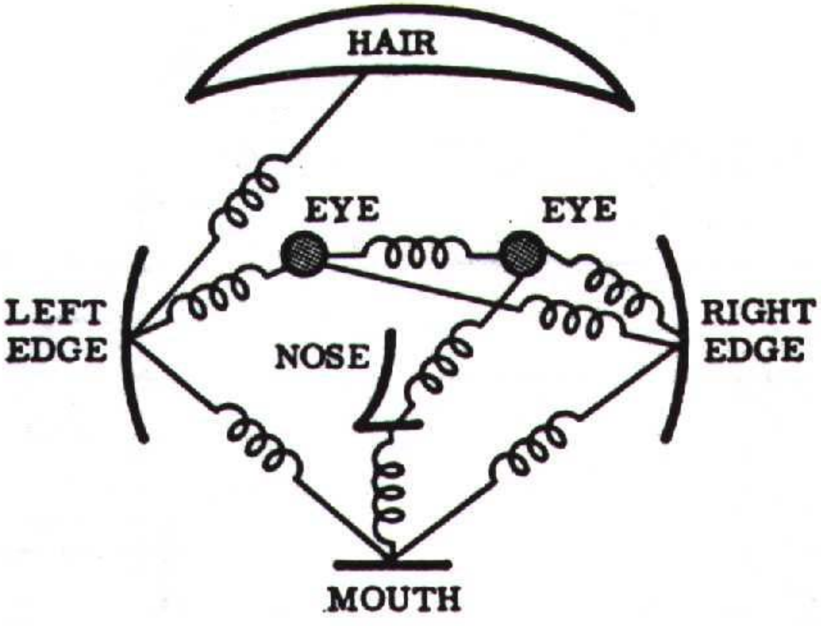
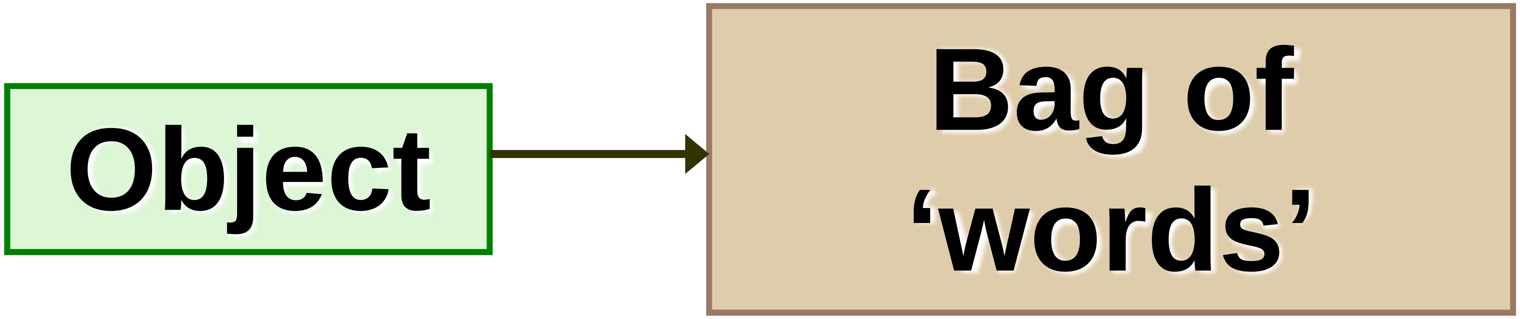


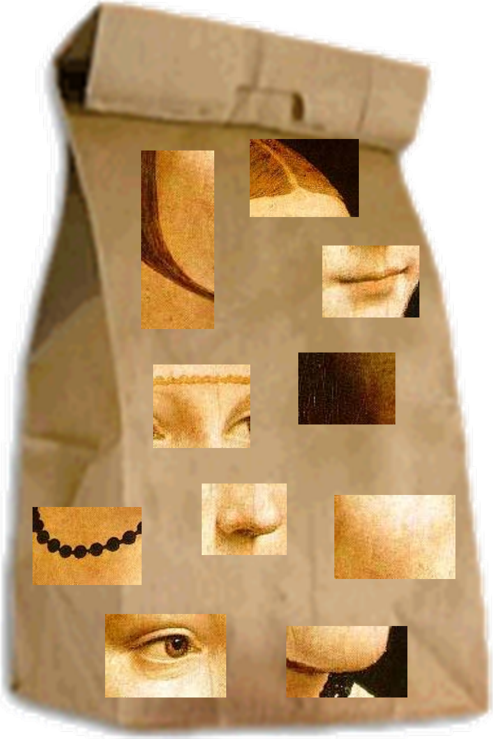
Figure from [Fischler & Elschlager 73]

Bag-of-features models



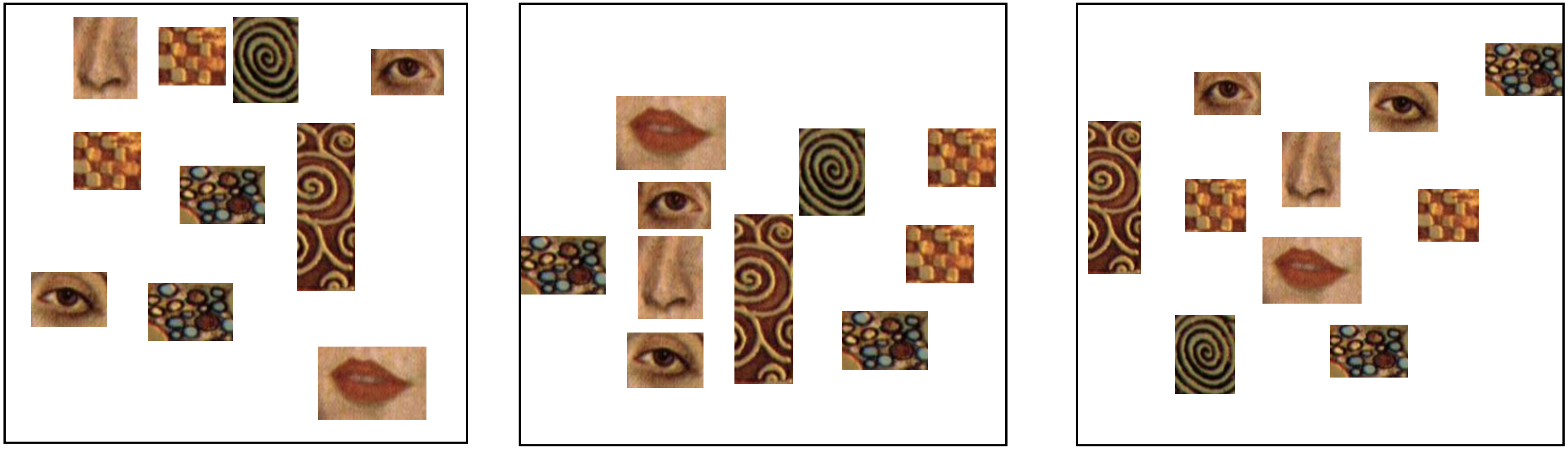
Object

Bag of ‘words’



Objects as texture

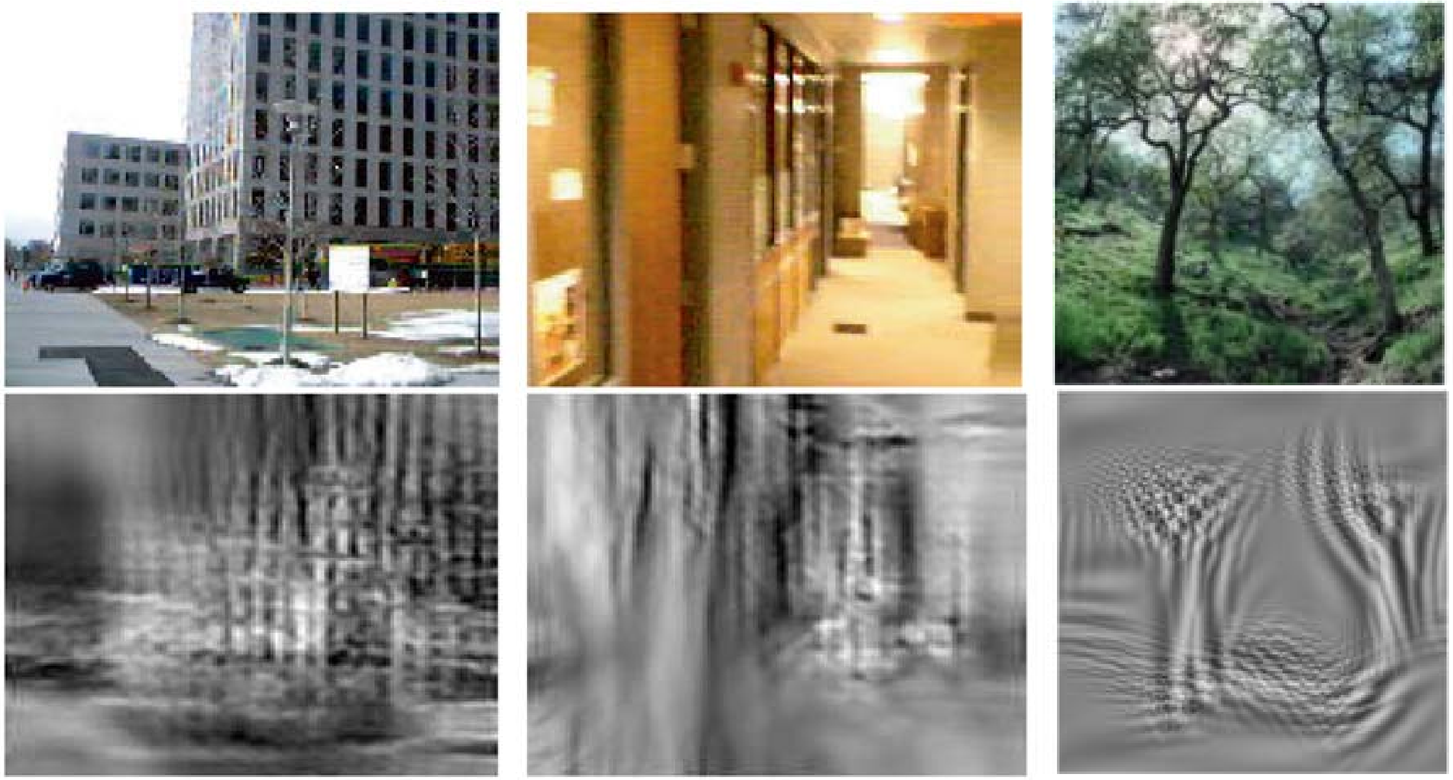
- All of these are treated as being the same



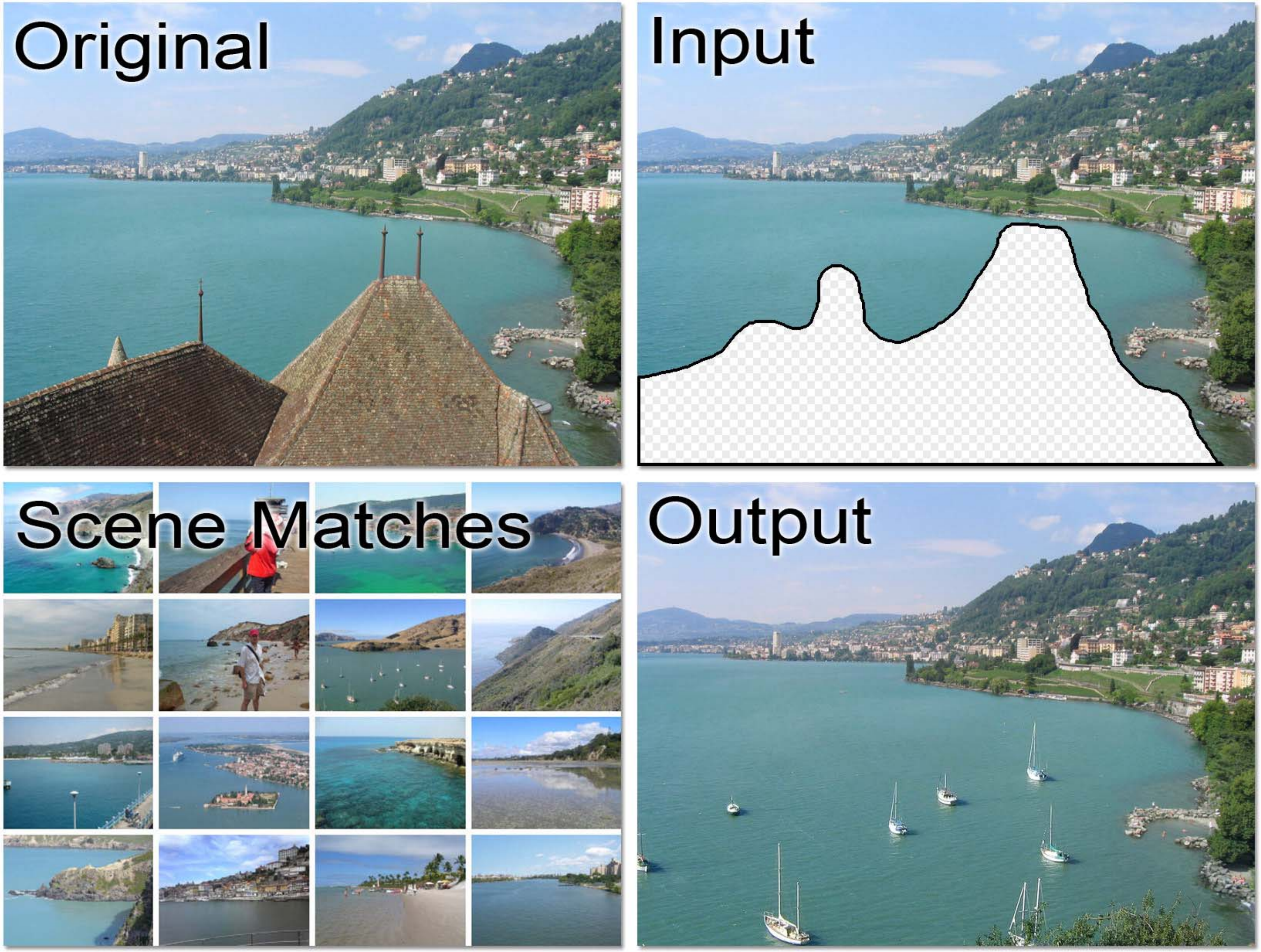
- No distinction between foreground and background: scene recognition?

Today: A comeback for global models?

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



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



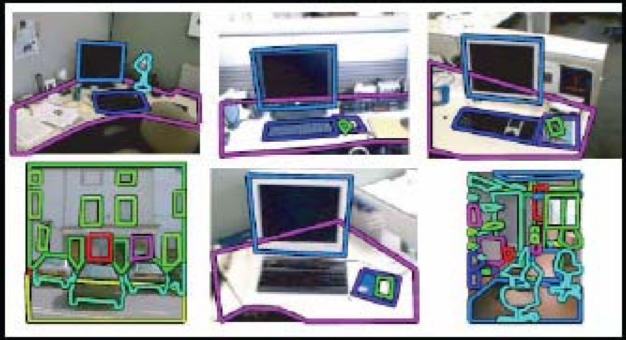
Object Recognition by Scene Alignment

Bryan C. Russell, Antonio Torralba, Ce Liu, Rob Fergus, William T. Freeman

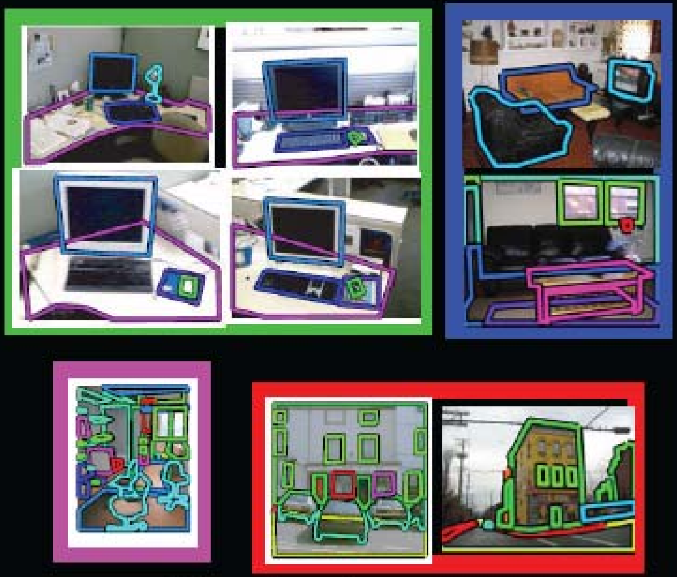
NIPS 2007

Goal: Recognize objects embedded in a scene

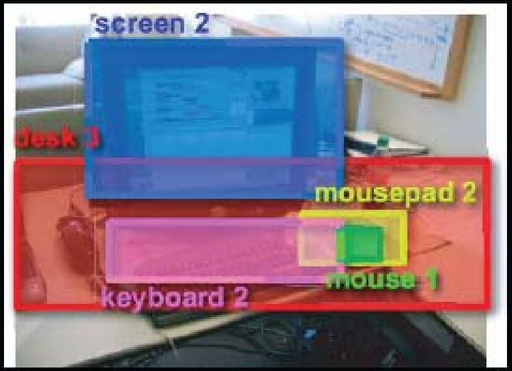
Input image



Nearest neighbors from 15,691 images



Cluster images using object labels



Output image with object labels transferred

Timeline of recognition

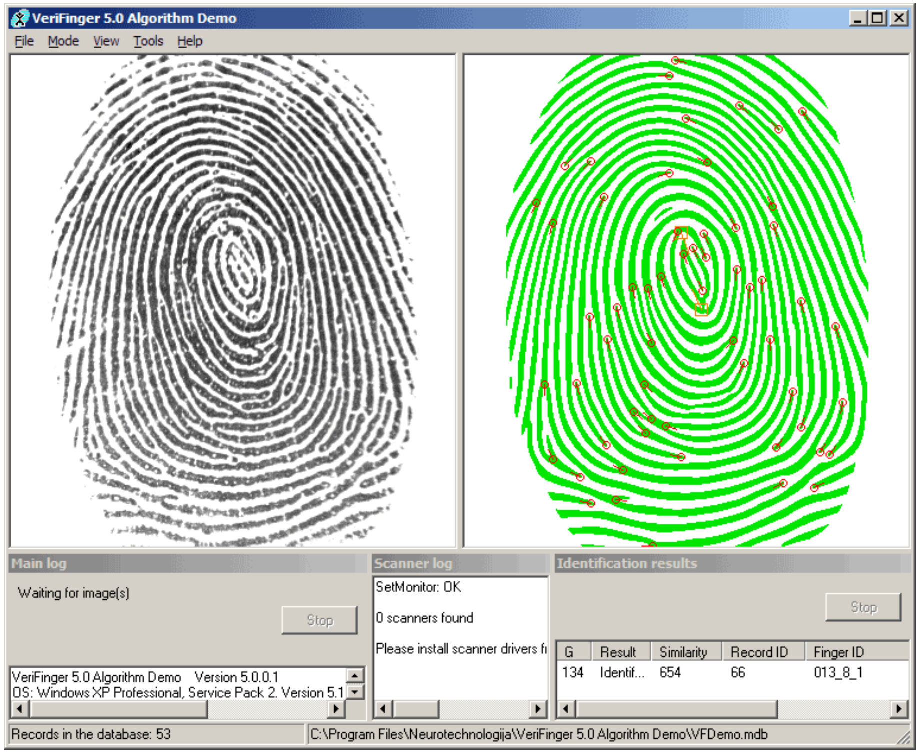
- 1965-late 1980s: alignment, geometric primitives
- Early 1990s: invariants, appearance-based methods
- Mid-late 1990s: sliding window approaches
- Late 1990s: feature-based methods
- Early 2000s: parts-and-shape models
- 2003 – present: bags of features
- Present trends: combination of local and global methods, modeling context, integrating recognition and segmentation

What “works” today

- Reading license plates, zip codes, checks

What “works” today

- Reading license plates, zip codes, checks
- Fingerprint recognition



Waiting for image(s)

```
VeriFinger 5.0 Algorithm Demo    Version 5.0.0.1
OS: Windows XP Professional, Service Pack 2. Version 5.1
```

```
SetMonitor: OK

0 scanners found

Please install scanner drivers f
```

G	Result	Similarity	Record ID	Finger ID
134	Identif...	654	66	013_8_1

Records in the database: 53

C:\Program Files\Neurotechnologija\VeriFinger 5.0 Algorithm Demo\VFDemo.mdb

What “works” today

- Reading license plates, zip codes, checks
- Fingerprint recognition
- Face detection

[Face priority AE] When a bright part of the face is too bright

What “works” today

- Reading license plates, zip codes, checks
- Fingerprint recognition
- Face detection
- Recognition of flat textured objects (CD covers, book covers, etc.)