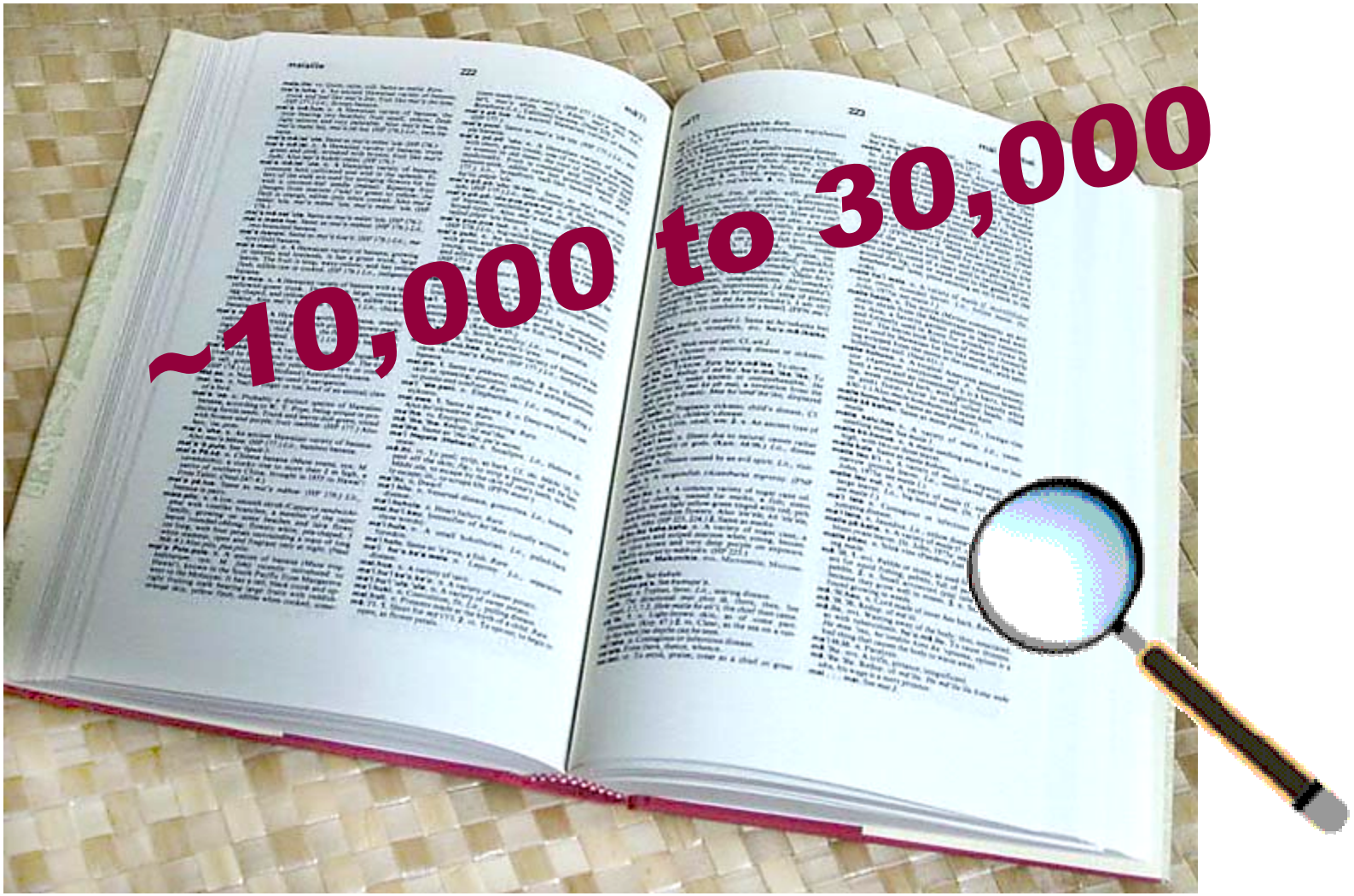
Recognition: Overview and History

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

How many visual object categories are there?



Biederman 1987


~10,000 to 30,000

OBJECTS

- ANIMALS
  - …..
  - VERTEBRATE
    - MAMMALS
      - TAPIR
      - BOAR
    - BIRDS
      - GROUSE
- PLANTS
- INANIMATE
  - NATURAL
  - MAN-MADE
    - CAMERA

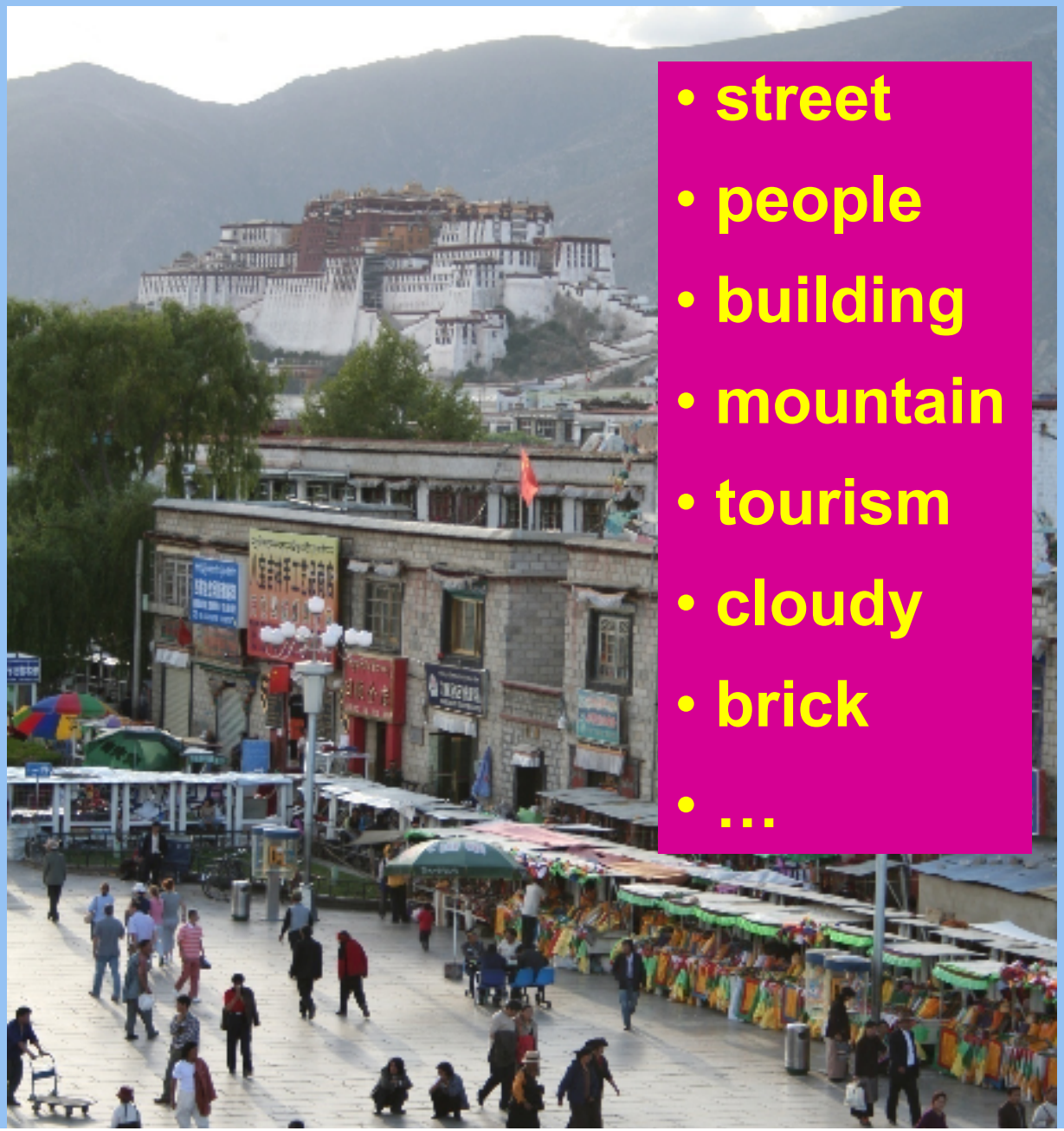
Specific recognition tasks

Scene categorization or classification

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

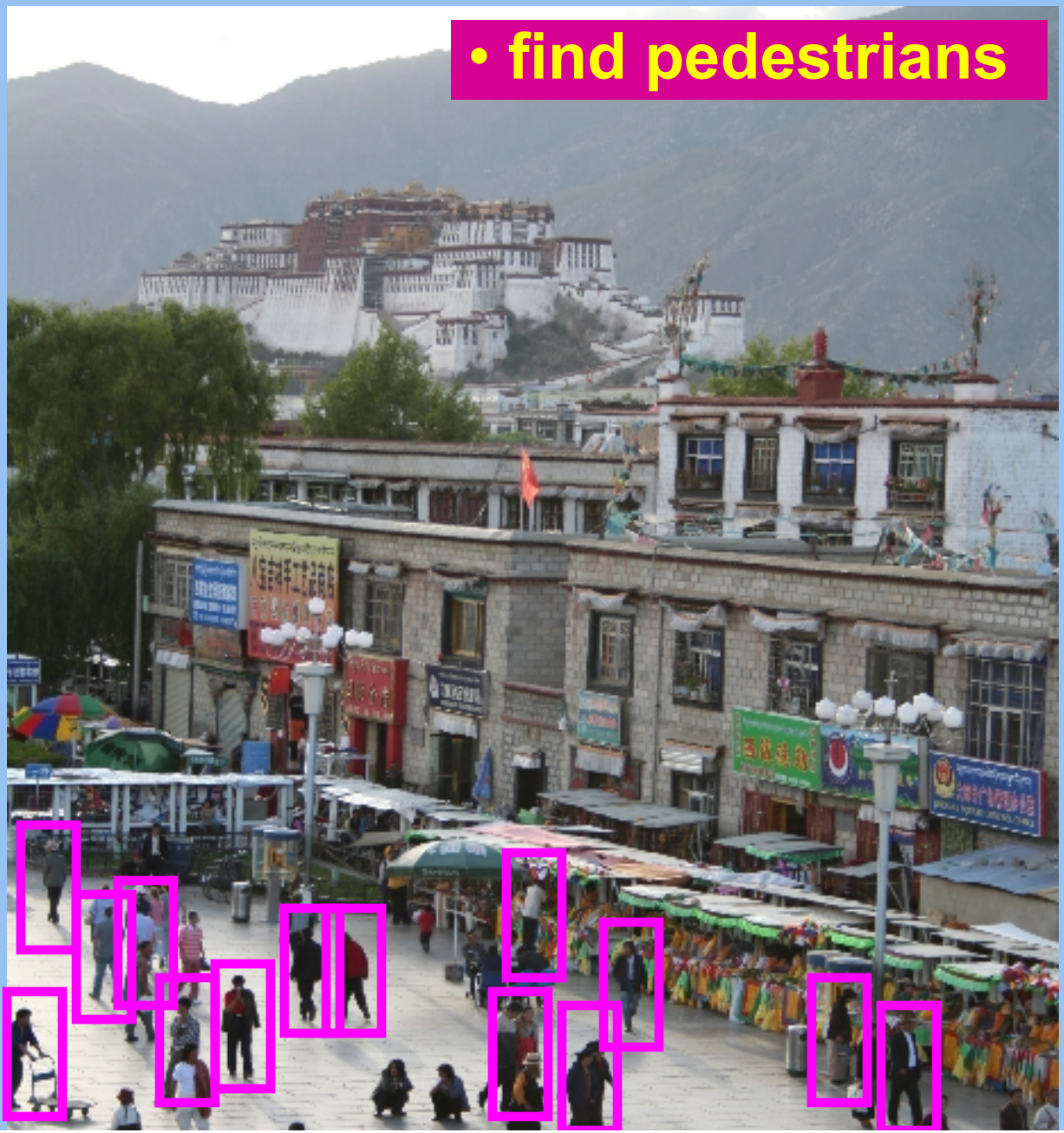
Svetlana Lazebnik

Image annotation / tagging / attributes



Svetlana Lazebnik

Object detection



Svetlana Lazebnik

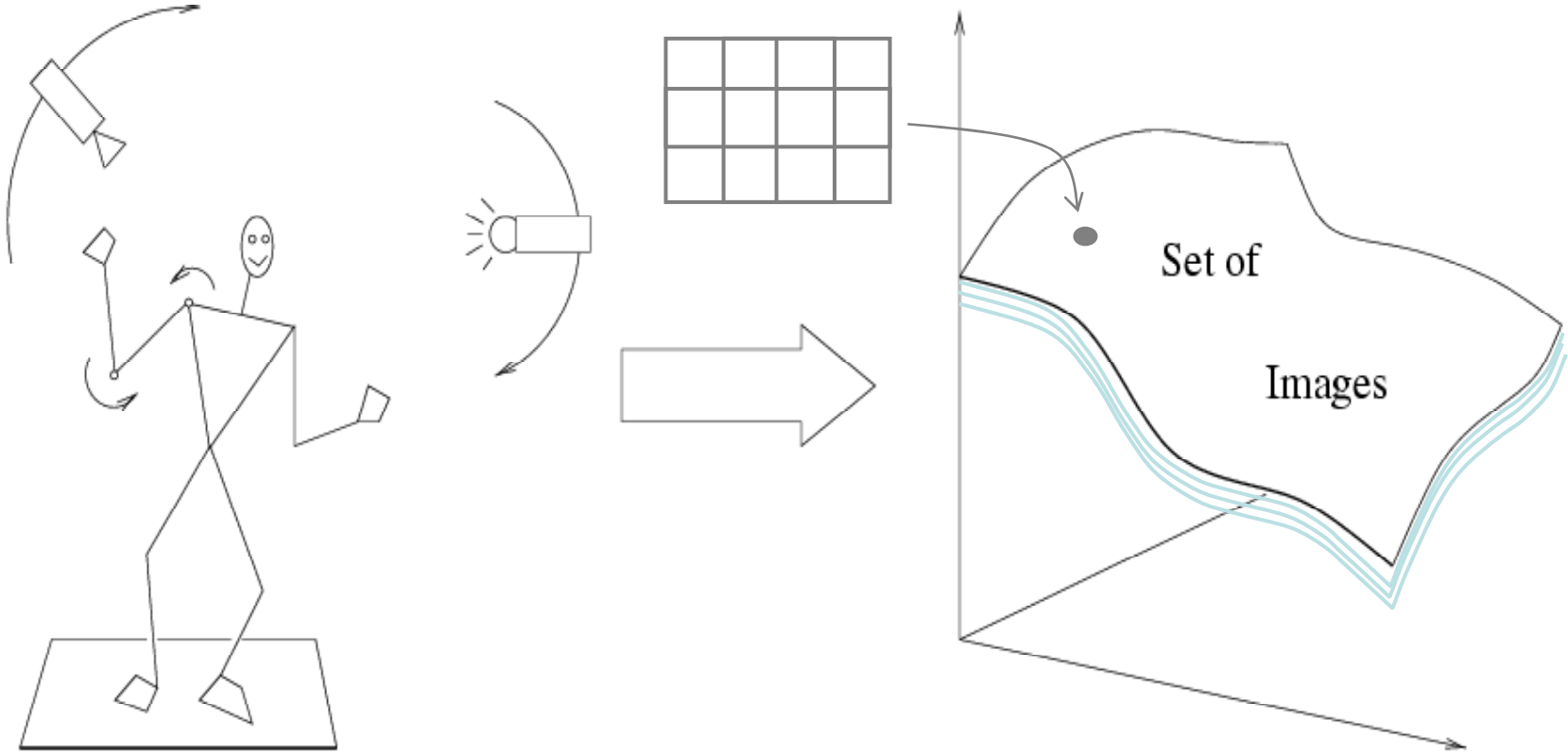
Image parsing

Svetlana Lazebnik

Scene understanding?

Svetlana Lazebnik

Recognition is all about modeling variability



Variability: Camera position
Illumination
Shape parameters

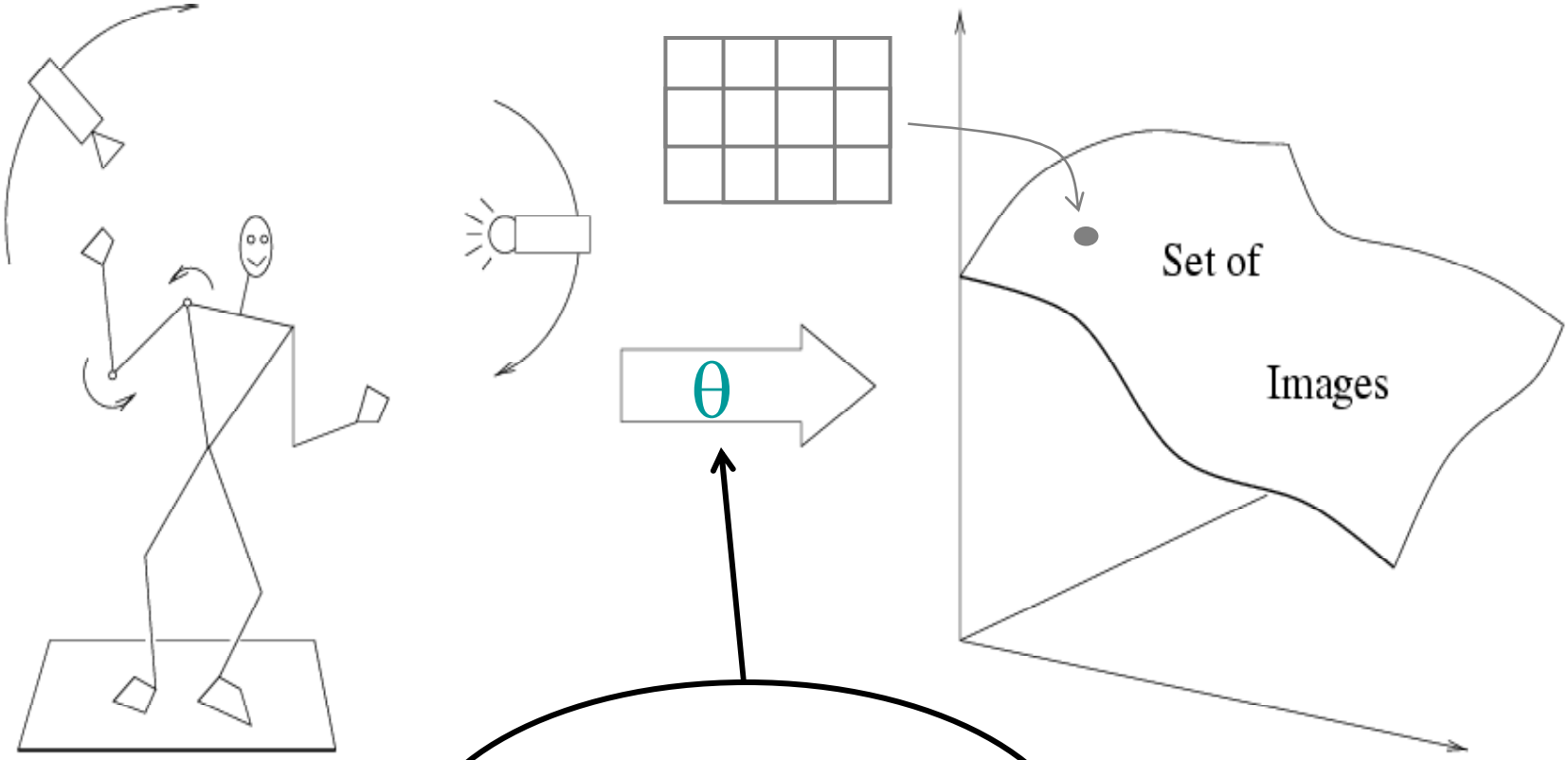
Within-class variations?

Within-class variations

Svetlana Lazebnik

History of ideas in recognition

- 1960s – early 1990s: the geometric era



Variability:

Camera position
Illumination

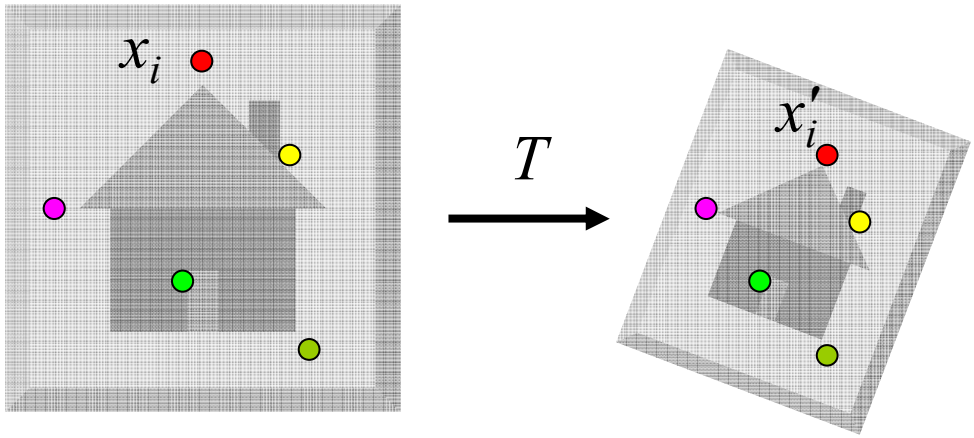
Alignment

Shape: assumed known

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

Recall: Alignment

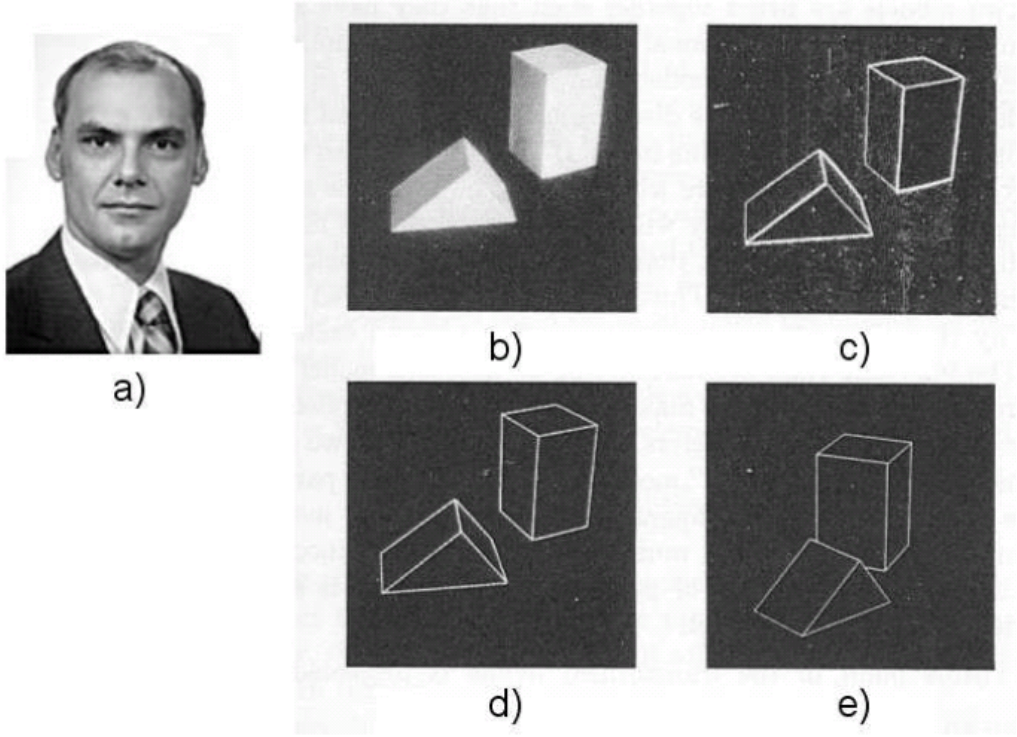
- Alignment: fitting a model to a transformation between pairs of features (*matches*) in two images



Find transformation T that minimizes

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

Recognition as an alignment problem: Block world

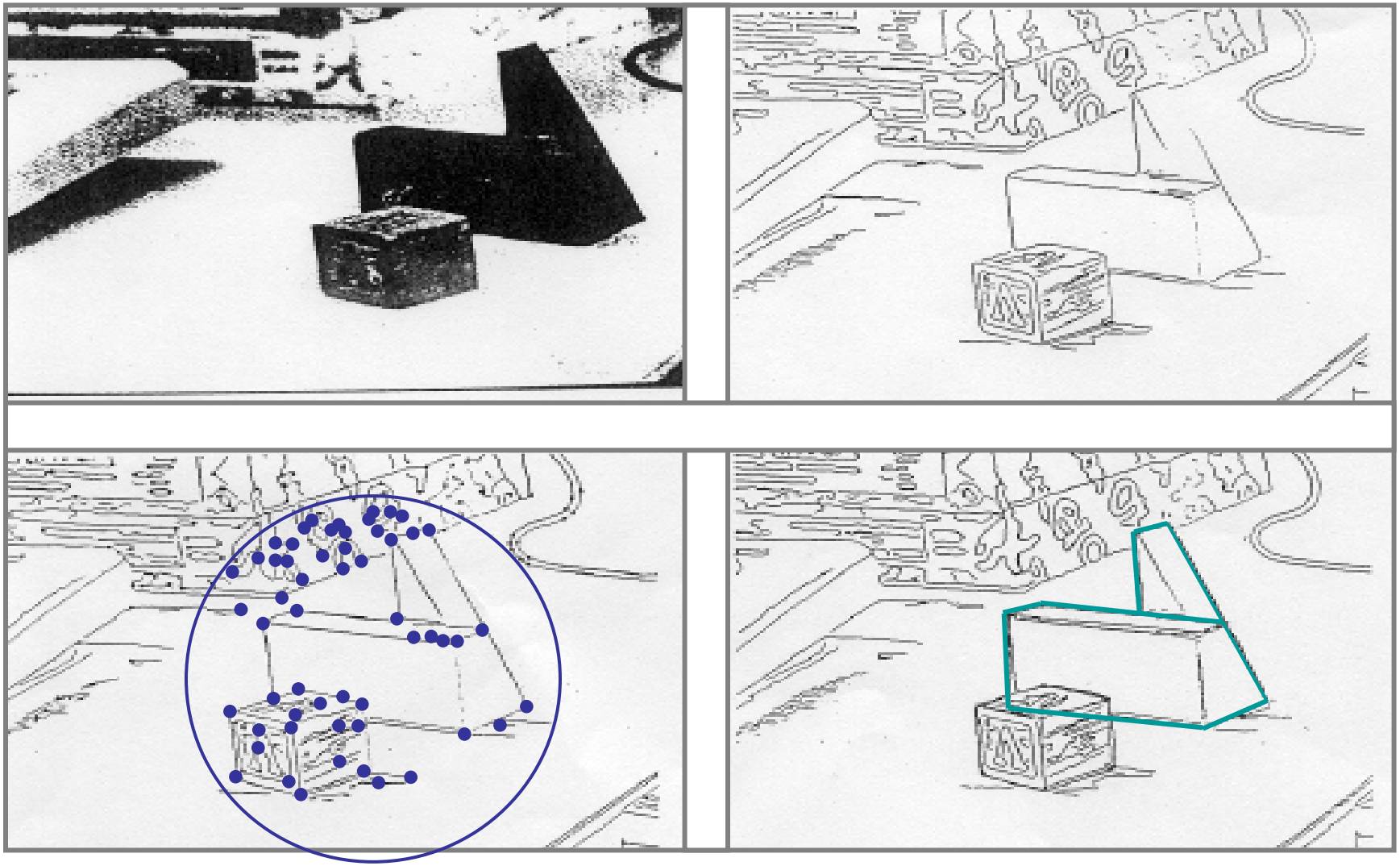


L. G. Roberts, *Machine Perception of Three Dimensional Solids*, Ph.D. thesis, MIT Department of Electrical Engineering, 1963.

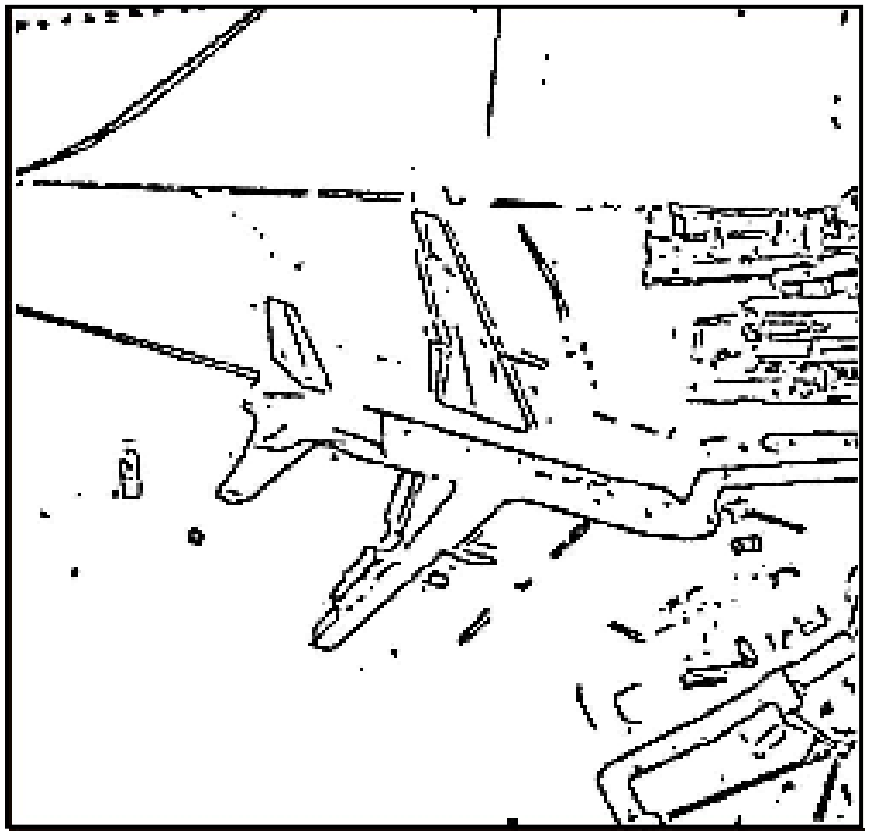
Fig. 1. A system for recognizing 3-d polyhedral scenes. a) L.G. Roberts. b)A blocks world scene. c)Detected edges using a 2x2 gradient operator. d) A 3-d polyhedral description of the scene, formed automatically from the single image. e) The 3-d scene displayed with a viewpoint different from the original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

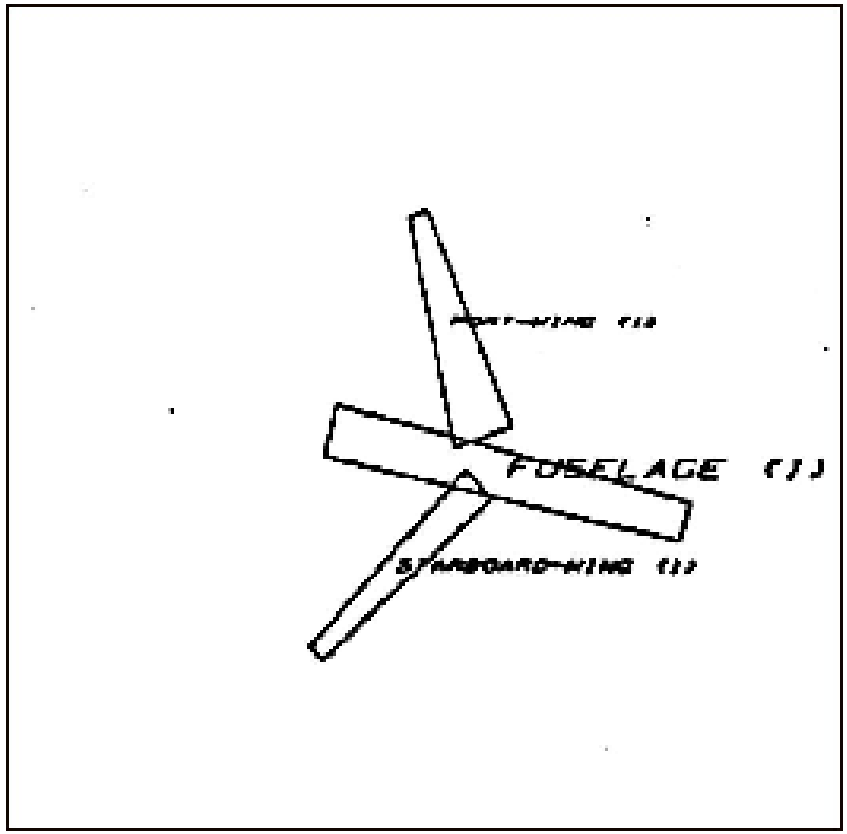
J. Mundy, Object Recognition in the Geometric Era: a Retrospective, 2006

Alignment: Huttenlocher & Ullman (1987)



Representing and recognizing object categories is harder...





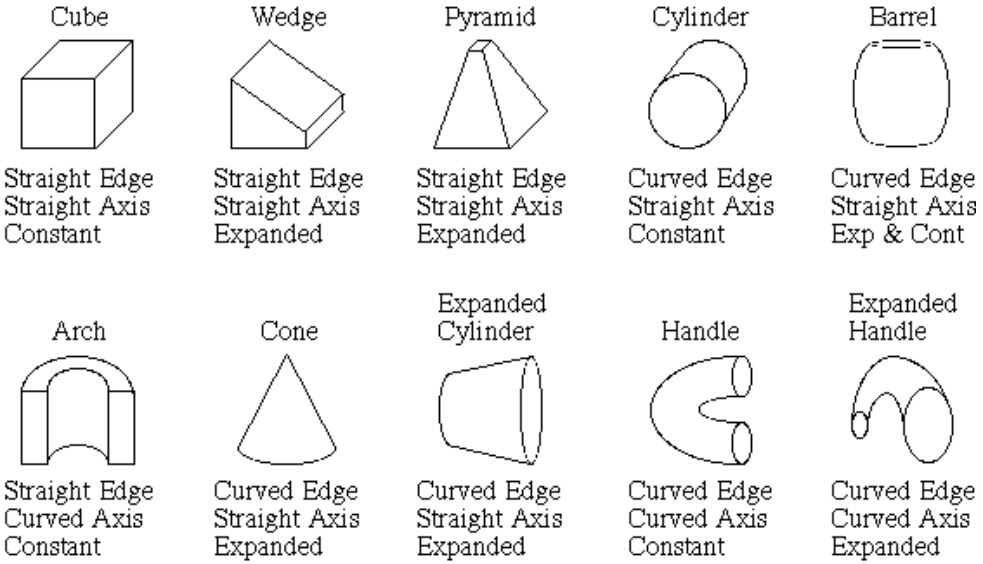
ACRONYM (Brooks and Binford, 1981)

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

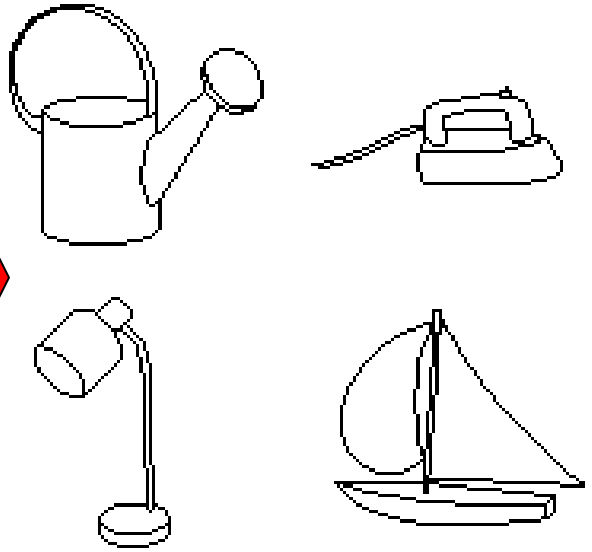
Recognition by components

Biederman (1987)

Primitives (geons)

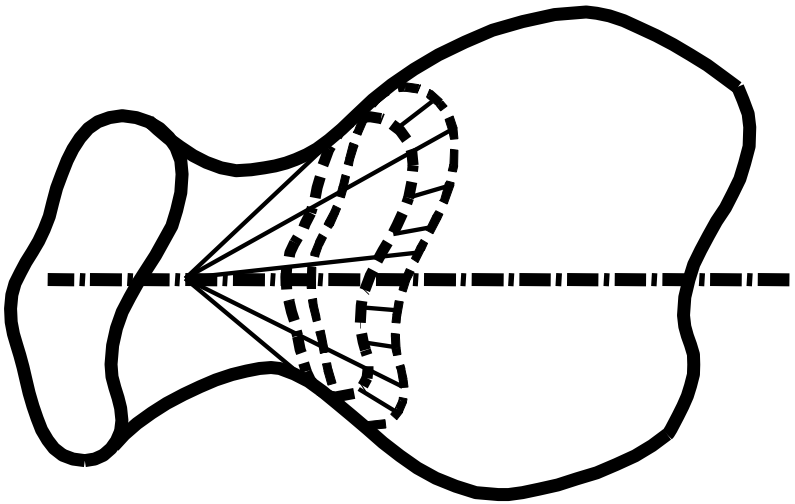


Objects

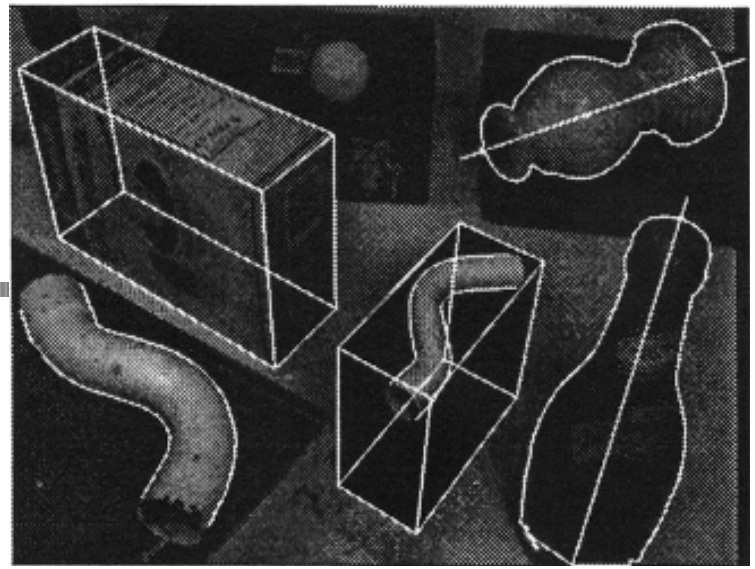


http://en.wikipedia.org/wiki/Recognition_by_Components_Theory

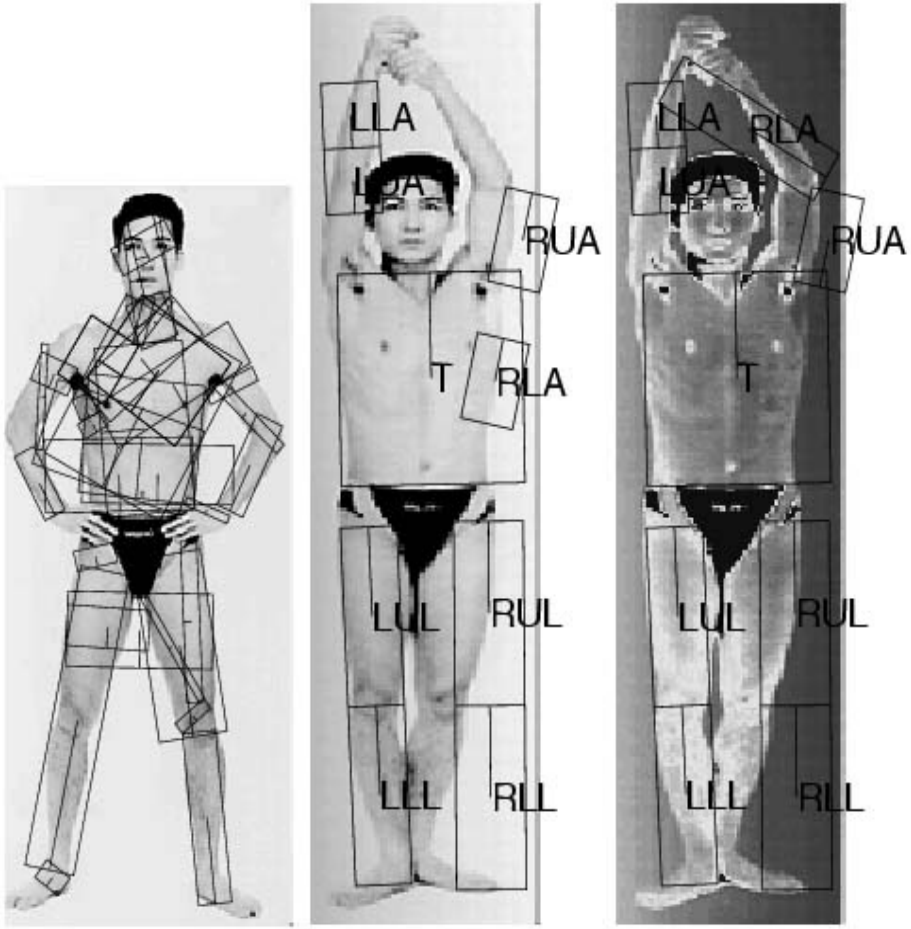
General shape primitives?



Generalized cylinders
Ponce et al. (1989)



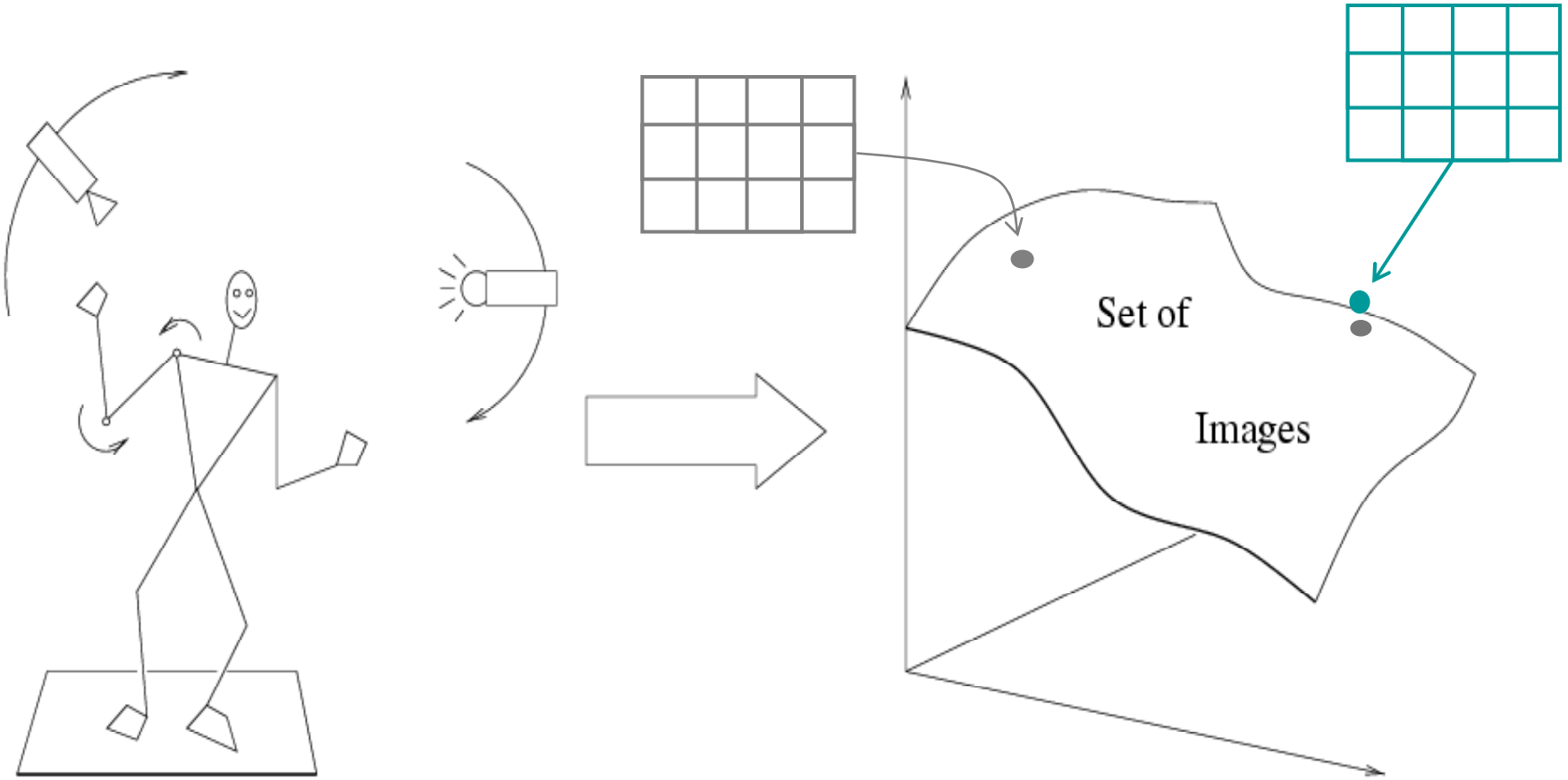
Zisserman et al. (1995)



Forsyth (2000)

History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models

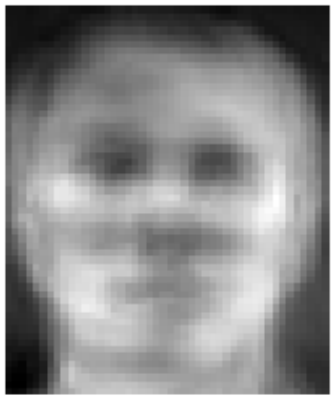


Empirical models of image variability

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

Svetlana Lazebnik

Limitations of global appearance models

- Requires global registration of patterns
- Not robust to clutter, occlusion, geometric transformations

Svetlana Lazebnik

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

- 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

History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- Mid-1990s: 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

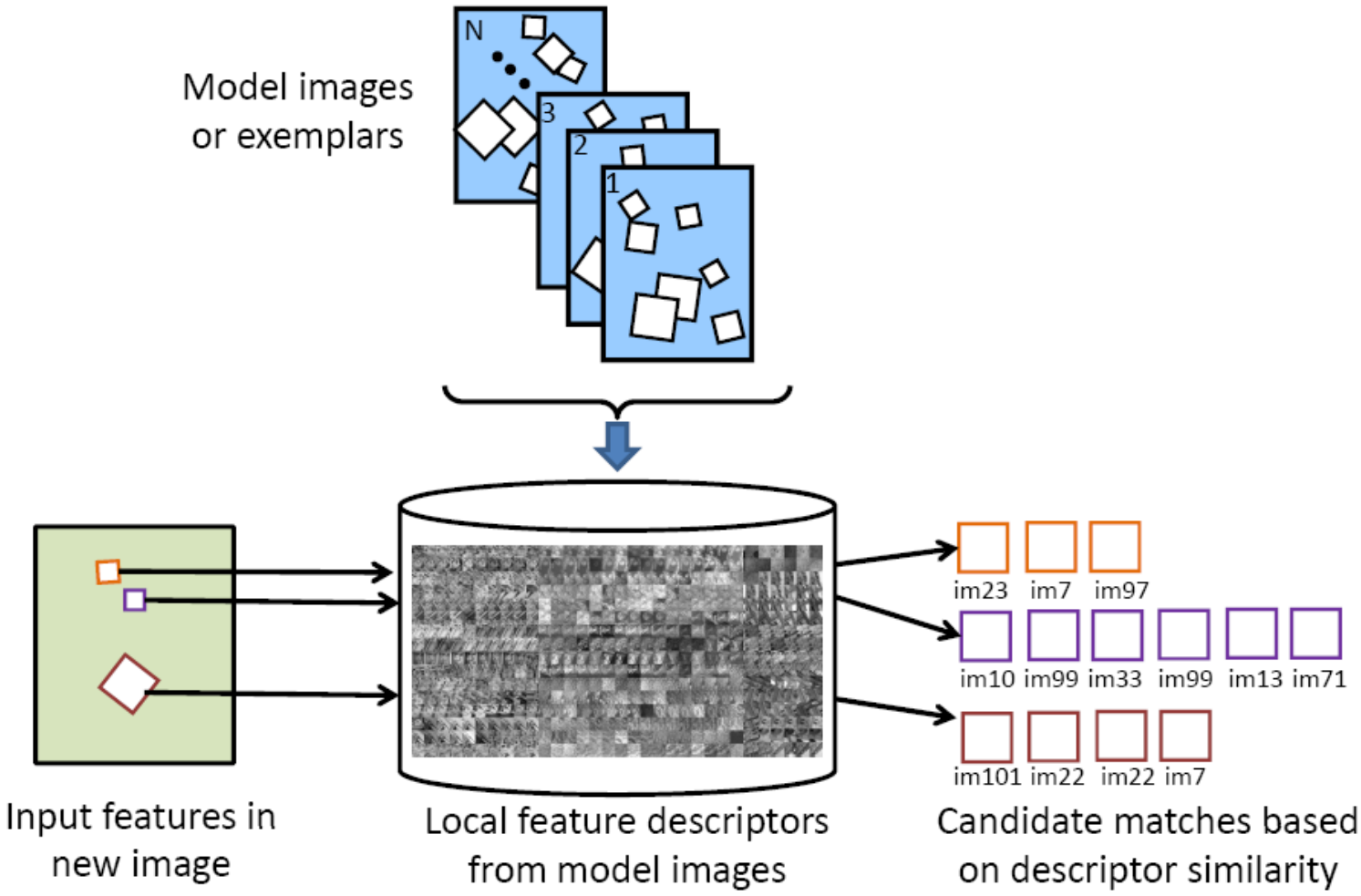


Image credit: K. Grauman and B. Leibe

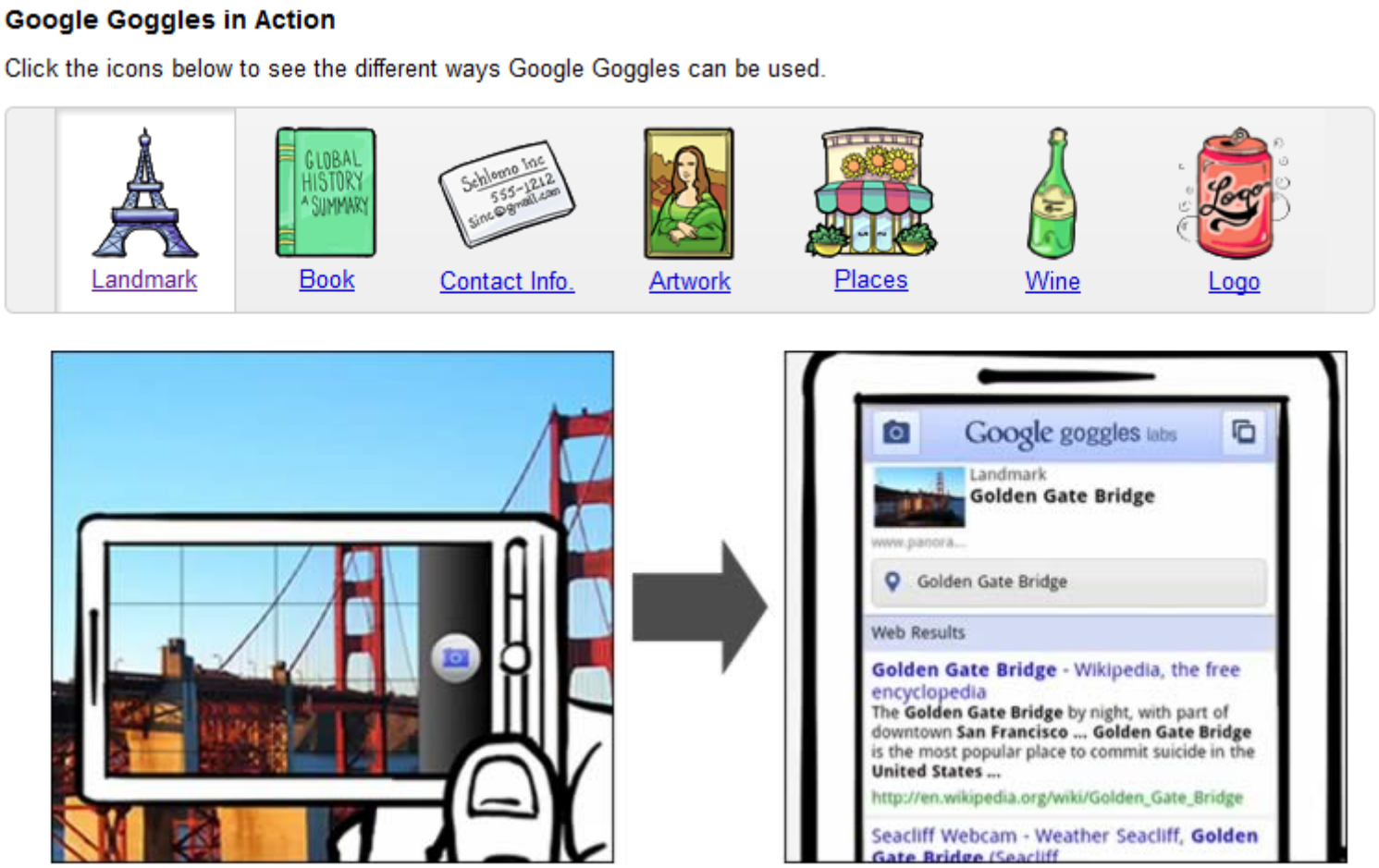
Large-scale image search

Combining local features, indexing, and spatial constraints

Philbin et al. ‘07

Large-scale image search

Combining local features, indexing, and spatial constraints



Available on phones that run Android 1.6+ (i.e. Donut or Eclair)

Svetlana Lazebnik

History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- Mid-1990s: 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

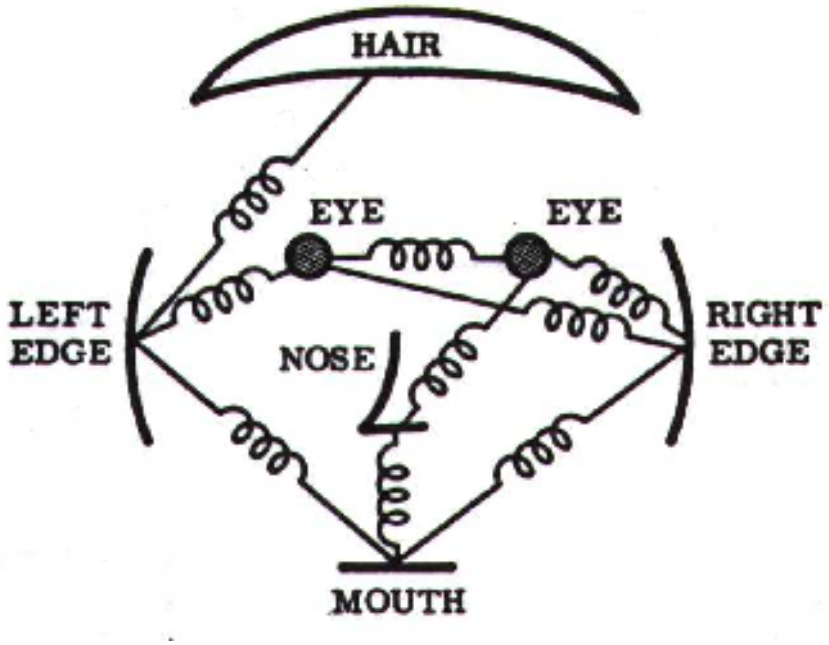


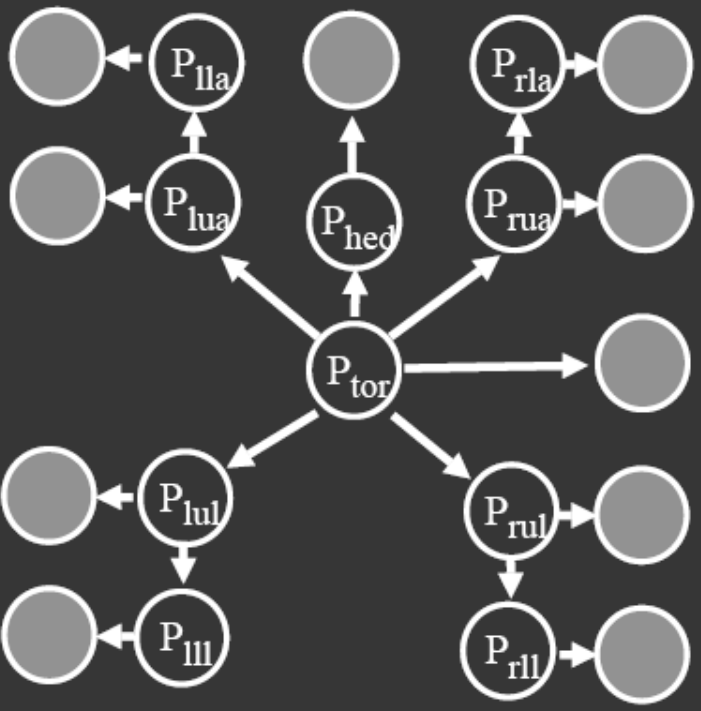
Figure from [Fischler & Elschlager 73]

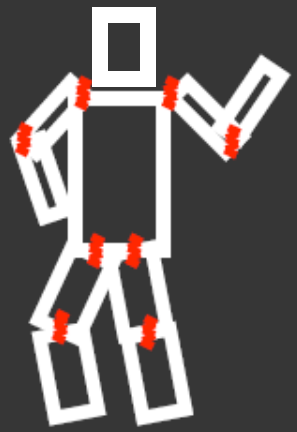
Constellation models

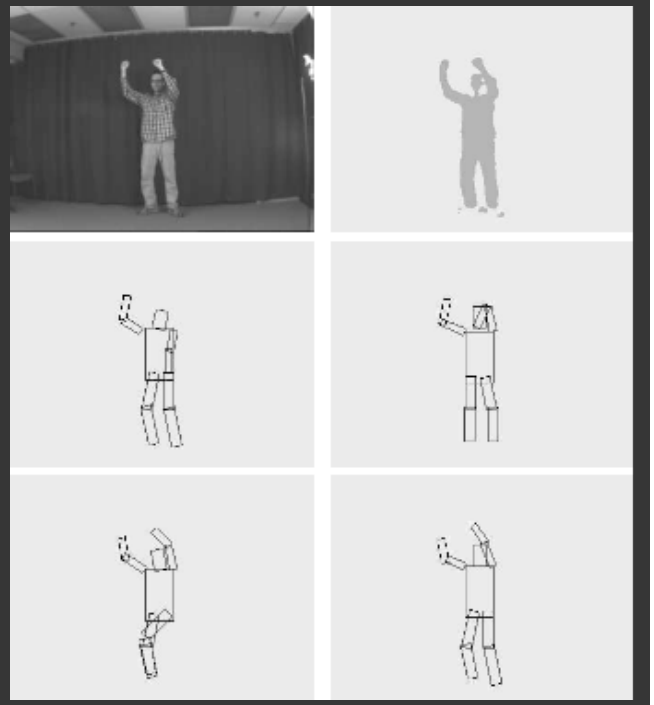
Weber, Welling & Perona (2000), Fergus, Perona & Zisserman (2003)

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
- Mid-1990s: sliding window approaches
- Late 1990s: local features
- Early 2000s: parts-and-shape models
- Mid-2000s: bags of features

Bag-of-features models

Object → **Bag of ‘words’**

Svetlana Lazebnik

Objects as texture

- All of these are treated as being the same

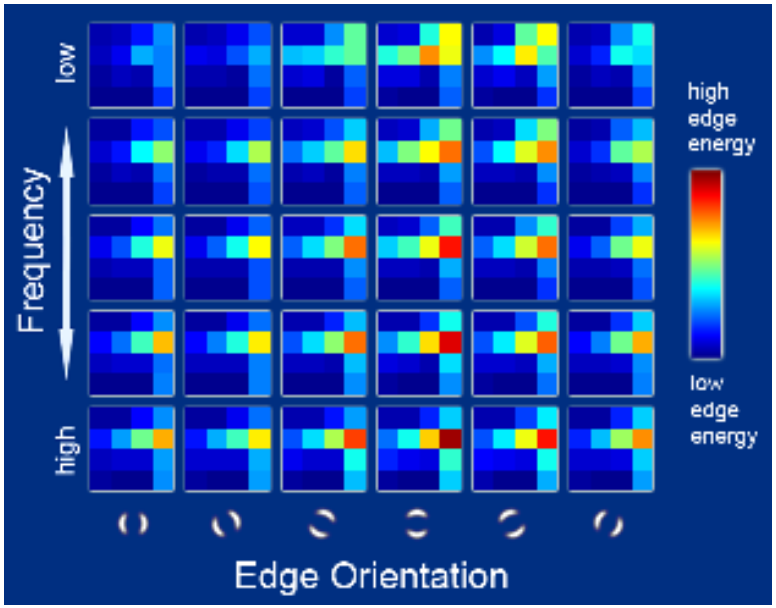
- No distinction between foreground and background: scene recognition?

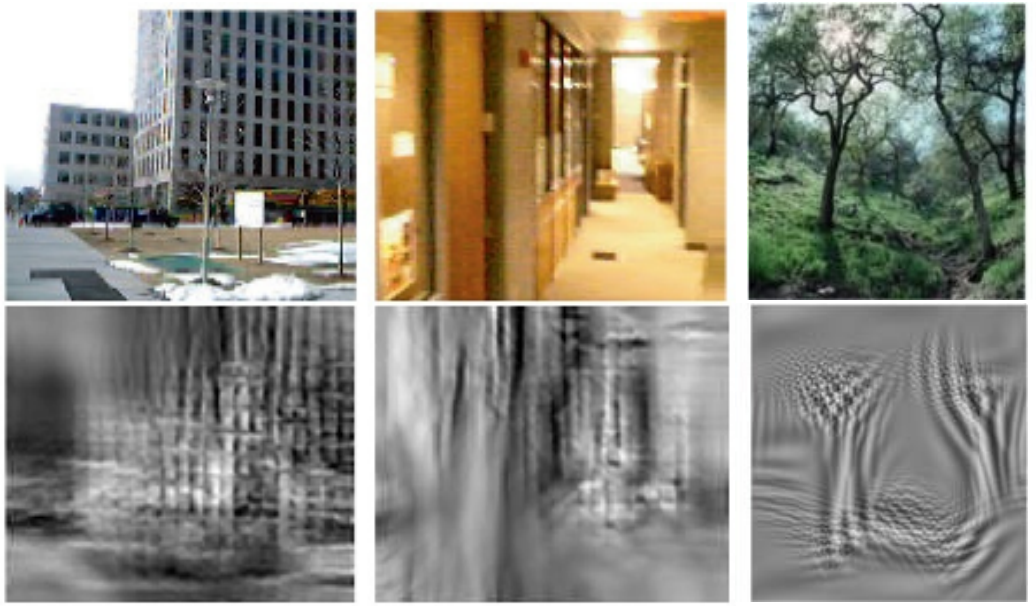
History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- Mid-1990s: sliding window approaches
- Late 1990s: local features
- Early 2000s: parts-and-shape models
- Mid-2000s: bags of features
- Present trends: combination of local and global methods, data-driven methods, context

Global scene descriptors

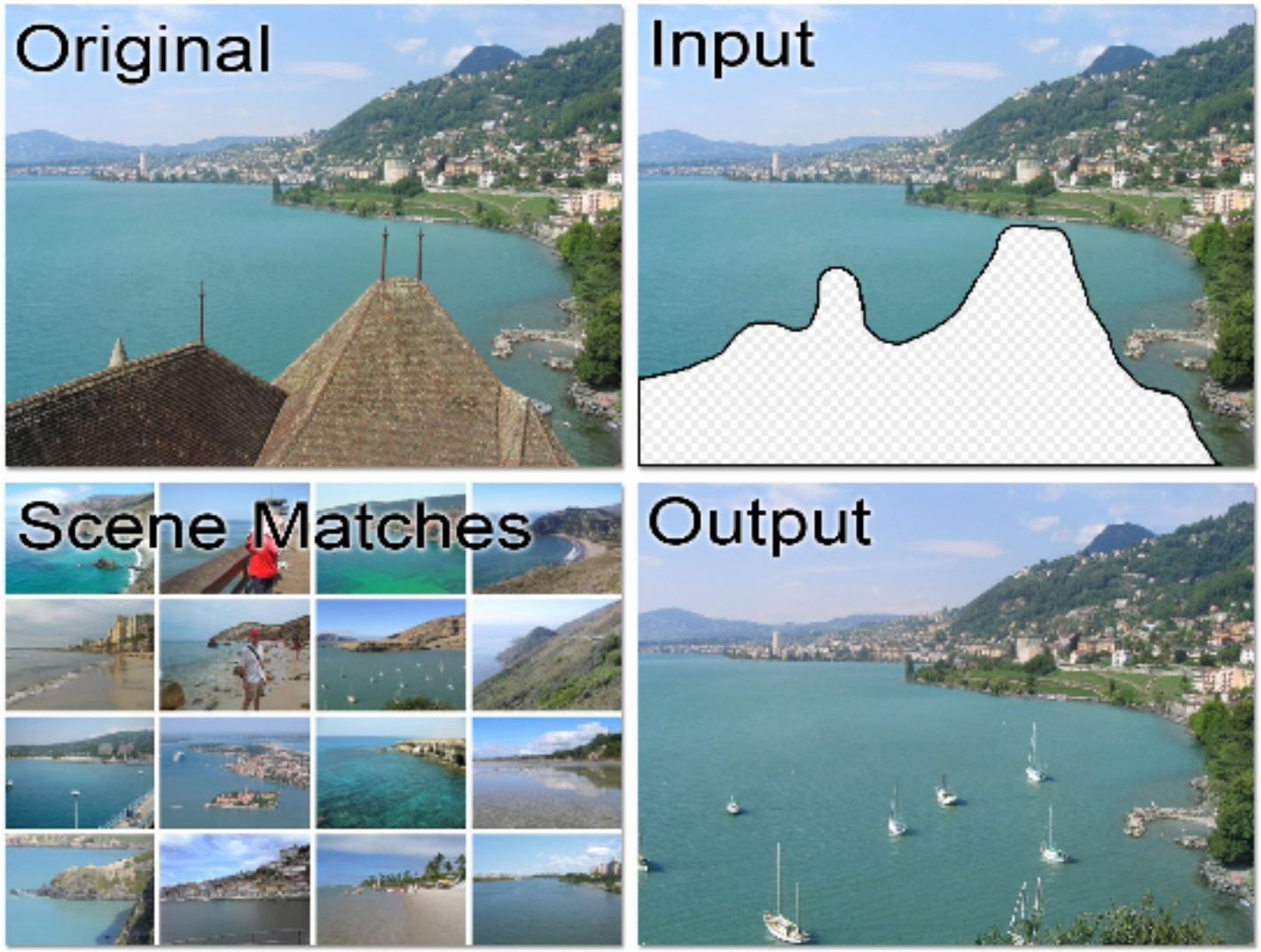
- The “gist” of a scene: Oliva & Torralba (2001)





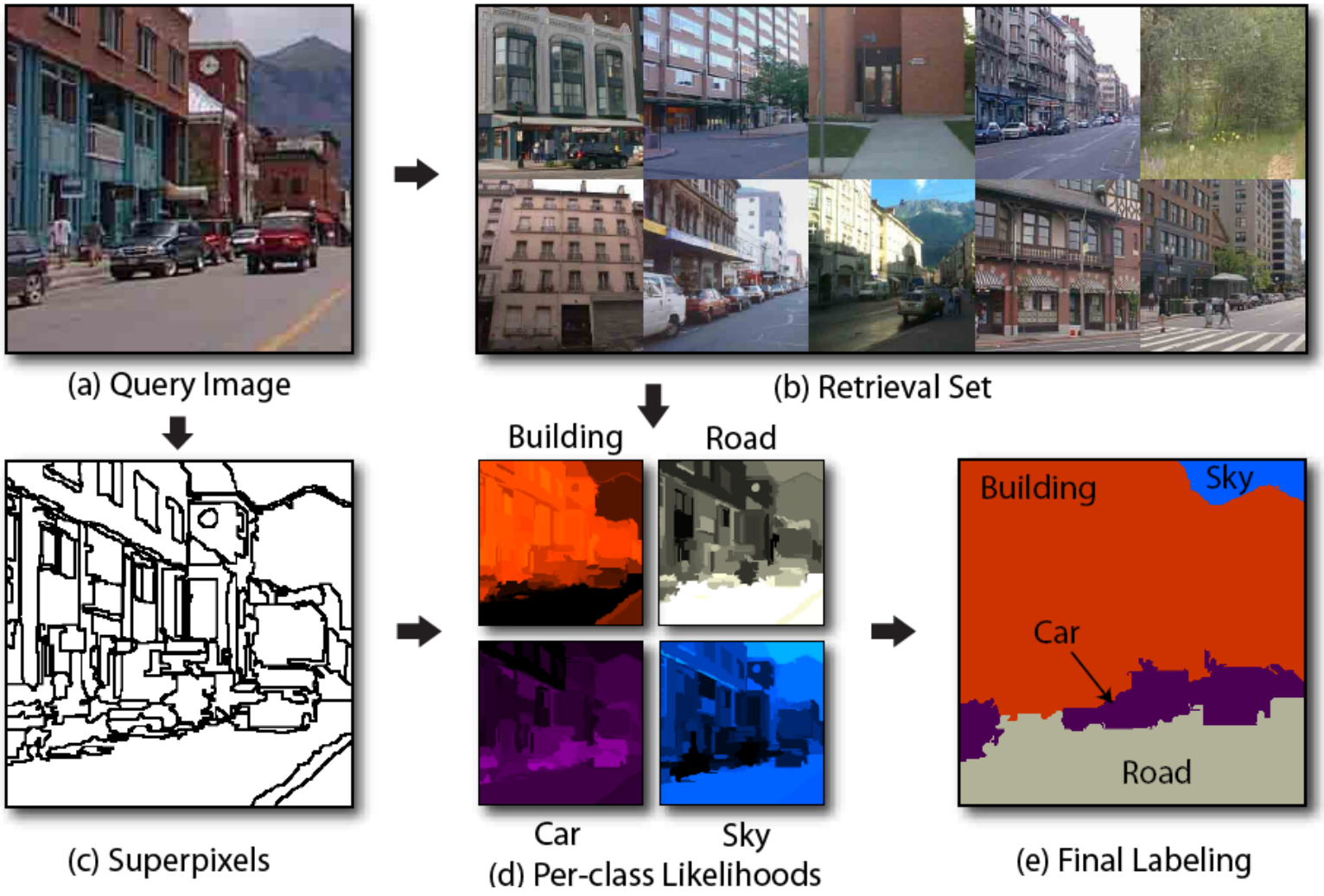
http://people.csail.mit.edu/torralba/code/spatialenvelope/

Data-driven methods



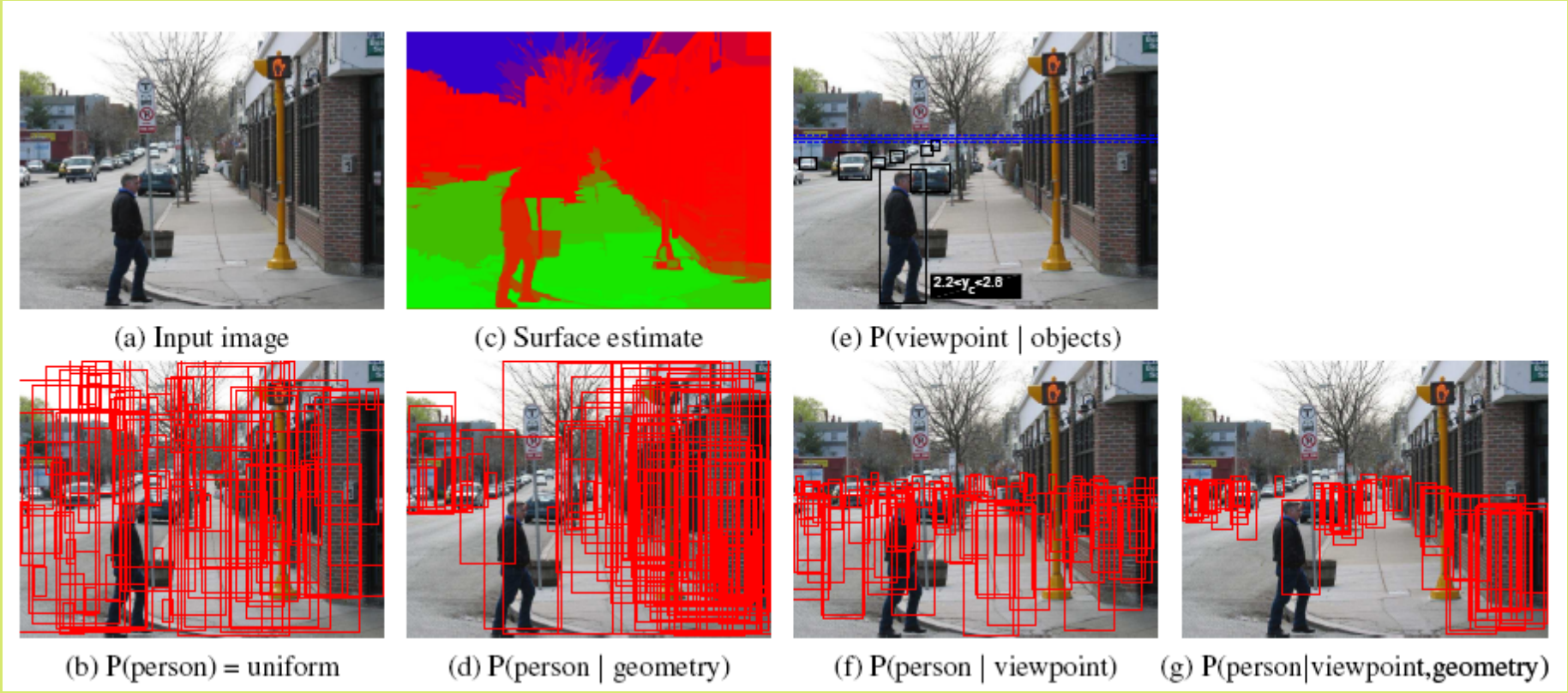
J. Hays and A. Efros, Scene Completion using Millions of Photographs, SIGGRAPH 2007

Data-driven methods



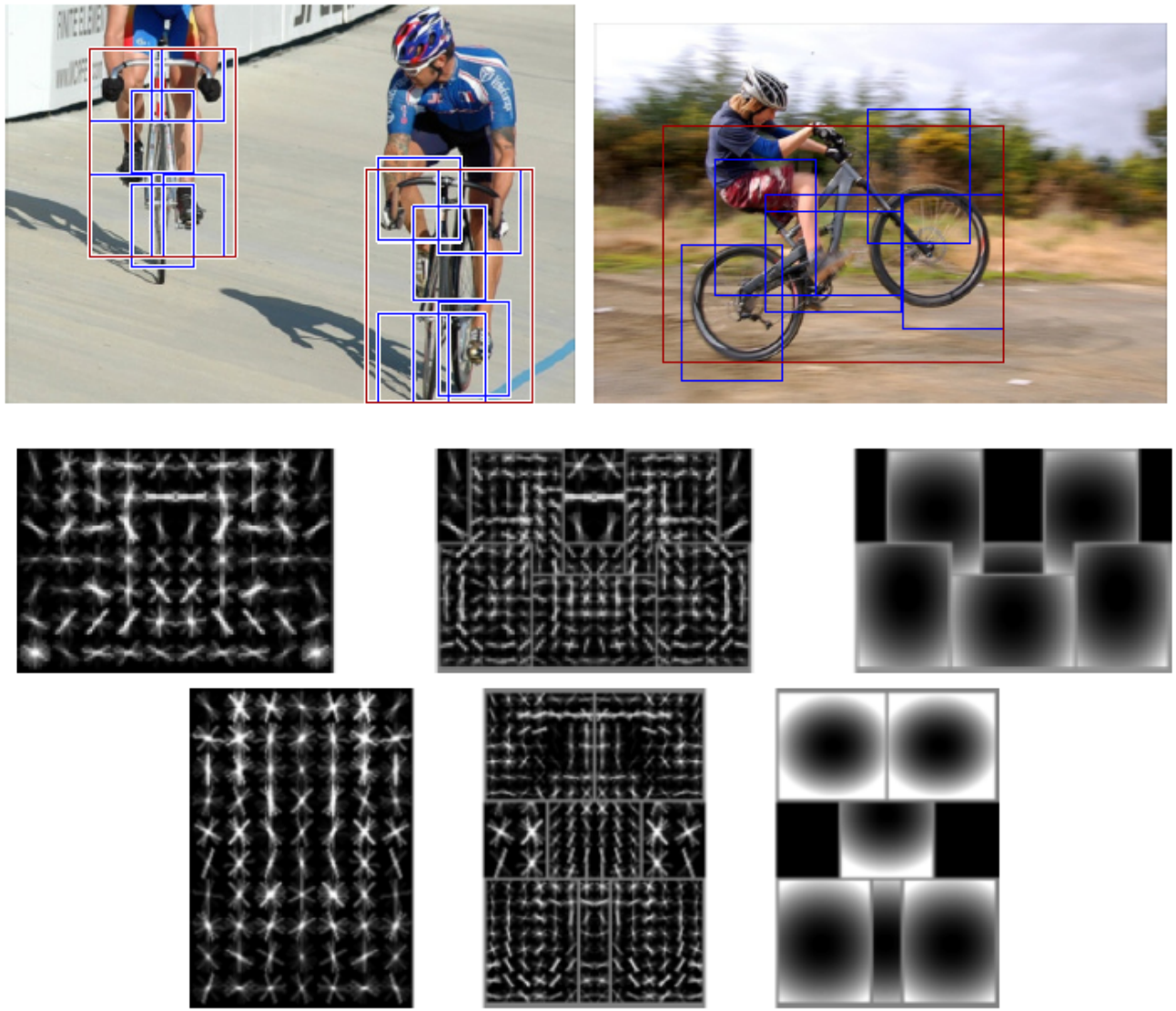
J. Tighe and S. Lazebnik, ECCV 2010

Geometric context



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

Discriminatively trained part-based models



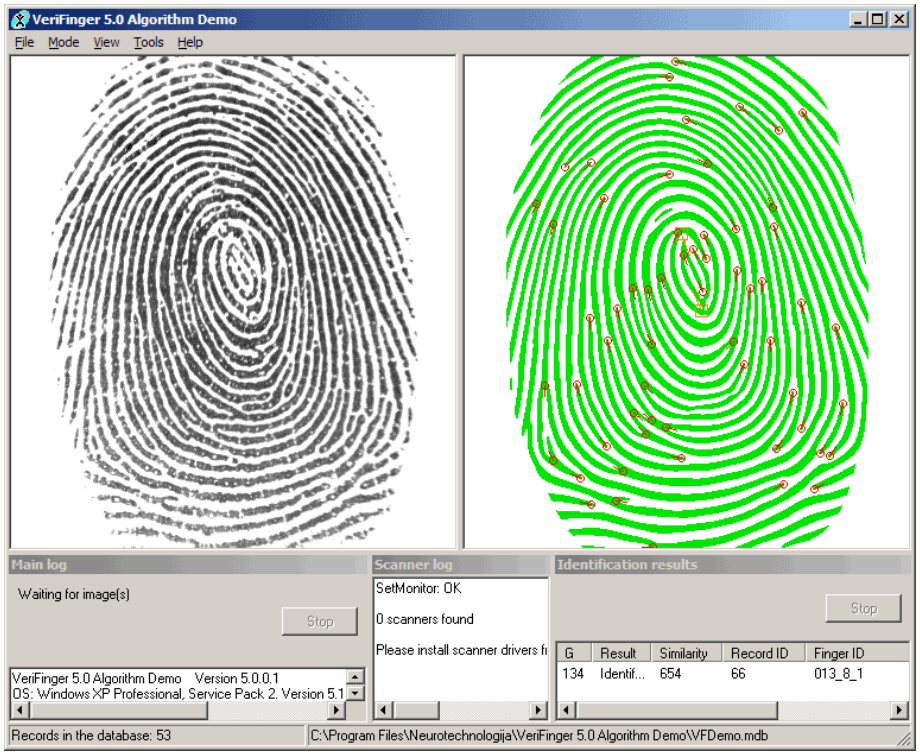
P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan, **"Object Detection with Discriminatively Trained Part-Based Models,"** PAMI 2009

What “works” today

- Reading license plates, zip codes, checks

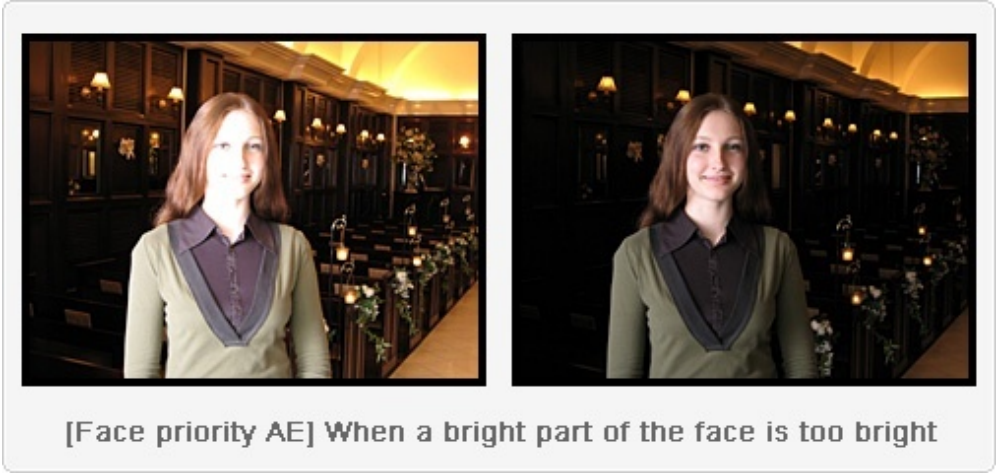
What “works” today

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



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.)