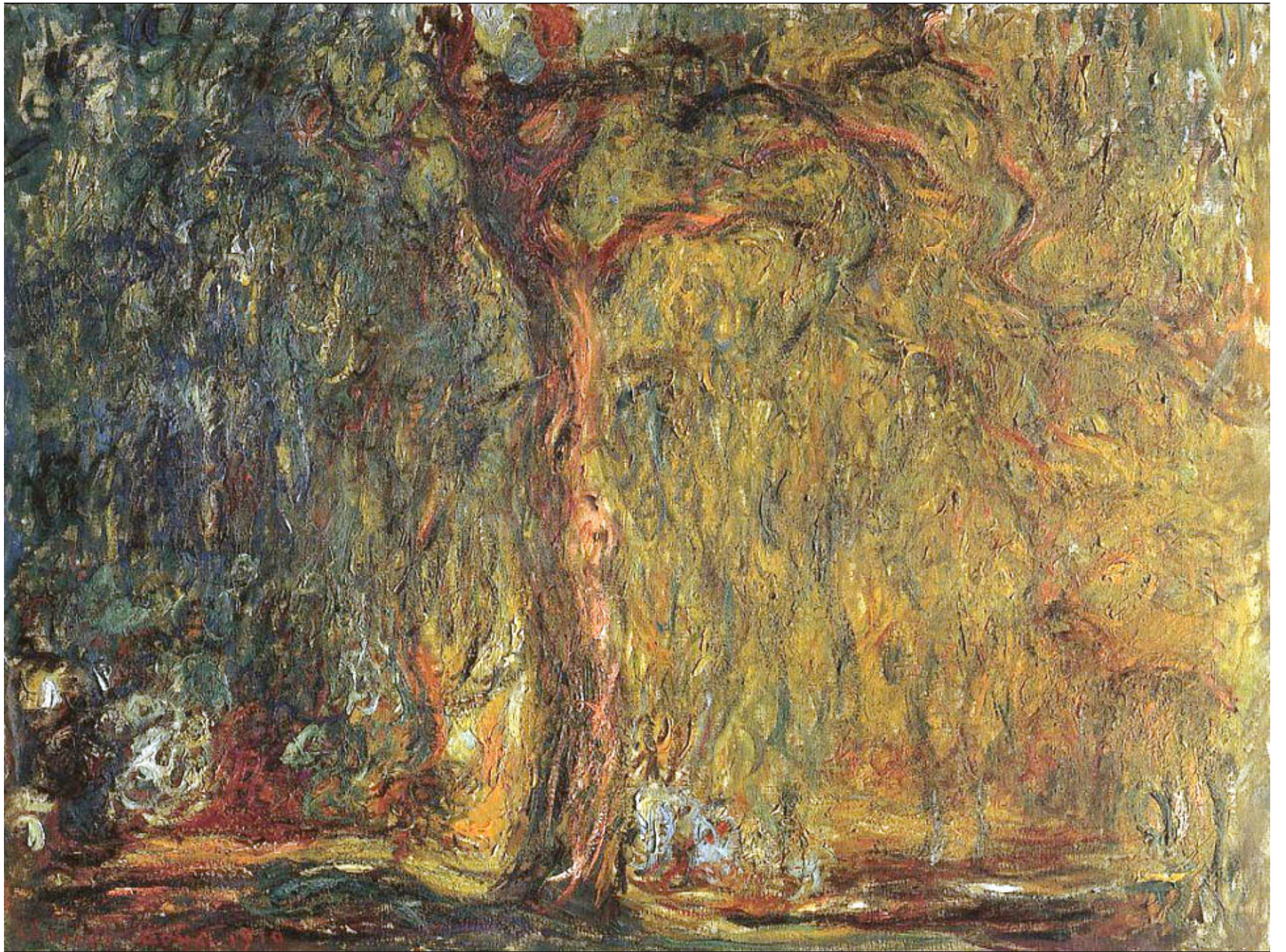




Reconnaissance d'objets et vision artificielle

<http://www.di.ens.fr/willow/teaching/recvis11/>

Jean Ponce (ponce@di.ens.fr)

<http://www.di.ens.fr/~ponce>

Equipe-projet WILLOW

ENS/INRIA/CNRS UMR 8548

Laboratoire d'Informatique

Ecole Normale Supérieure, Paris

Jean Ponce



<http://www.di.ens.fr/~ponce/>

Cordelia Schmid



<http://lear.inrialpes.fr/~schmid/>

Josef Sivic



<http://www.di.ens.fr/~josef/>

Ivan Laptev



<http://www.irisa.fr/vista/Equipe/People/Ivan.Laptev.html>

Nous cherchons
toujours
des stagiaires
à la fin du cours

Initiation à la vision artificielle

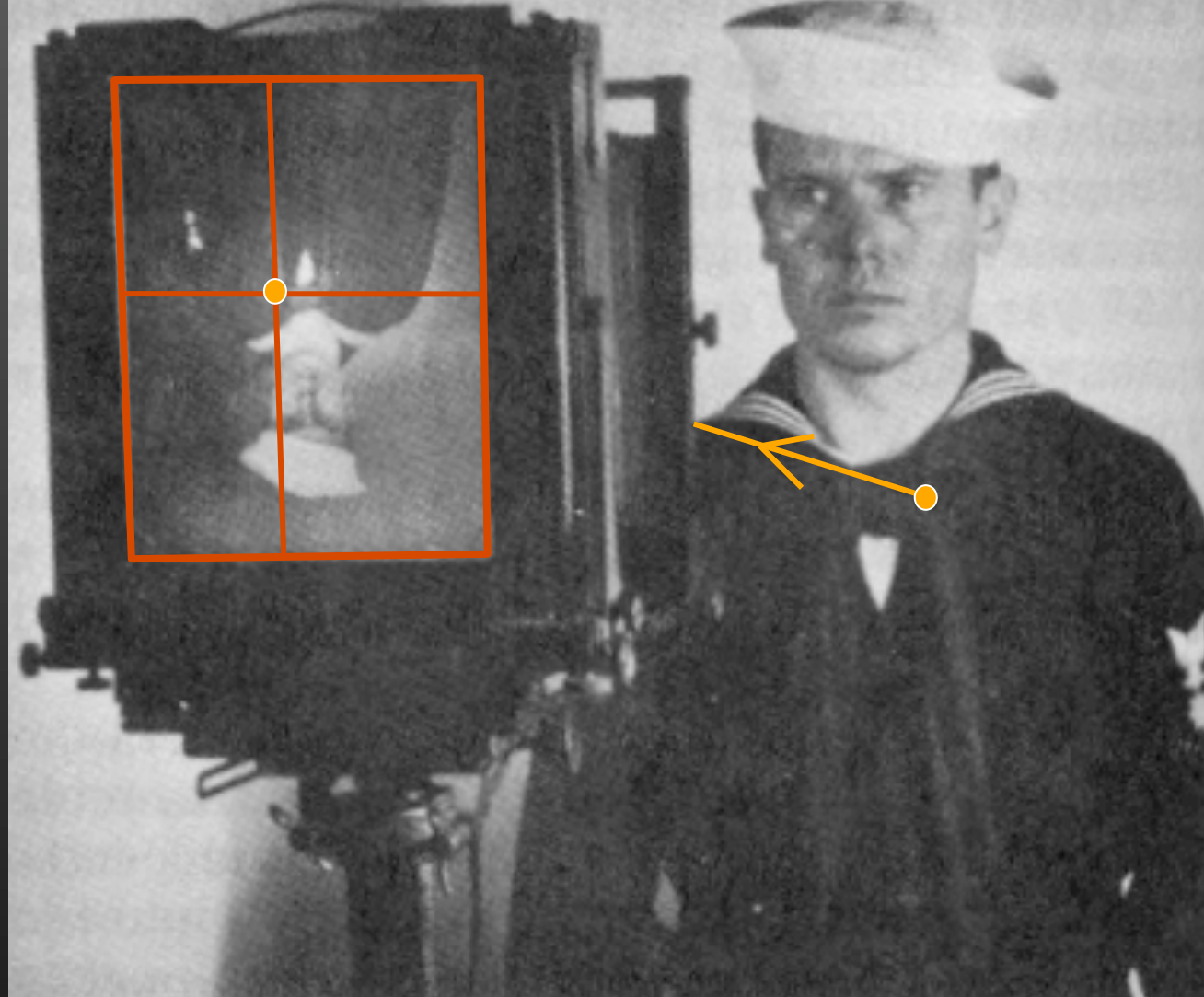
Jean Ponce (ponce@di.ens.fr)

Jeudis, salle U/V, 9-12h

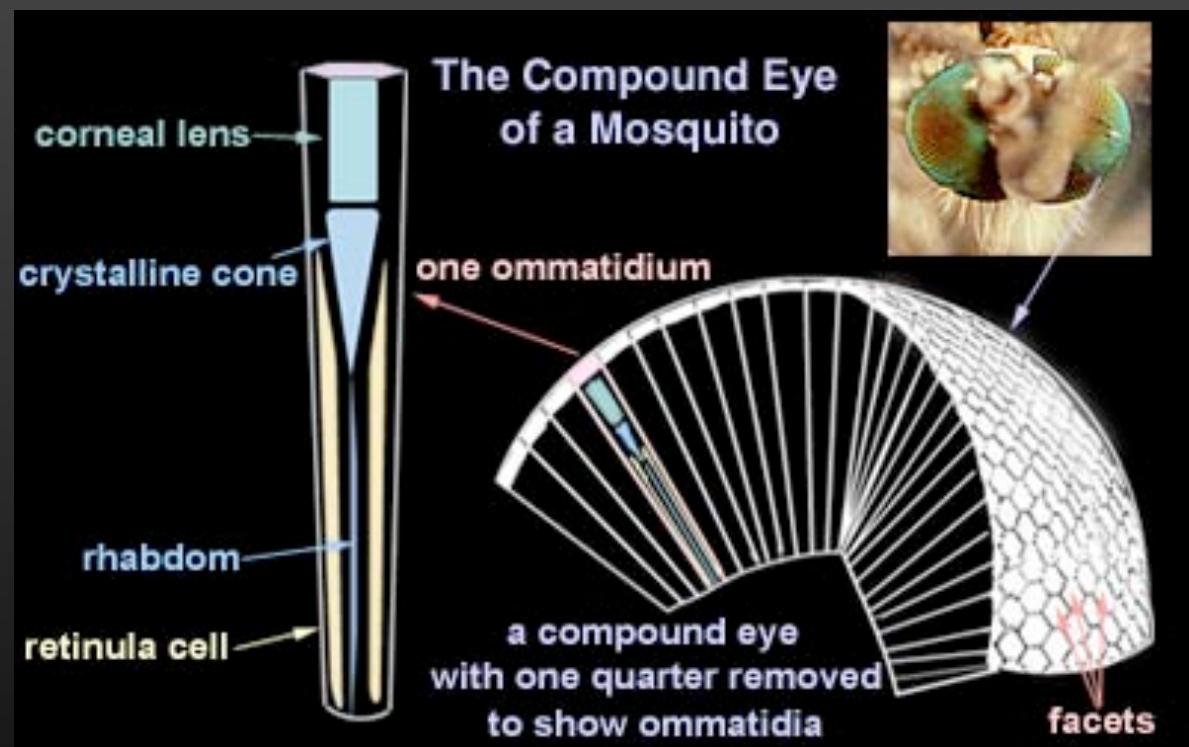
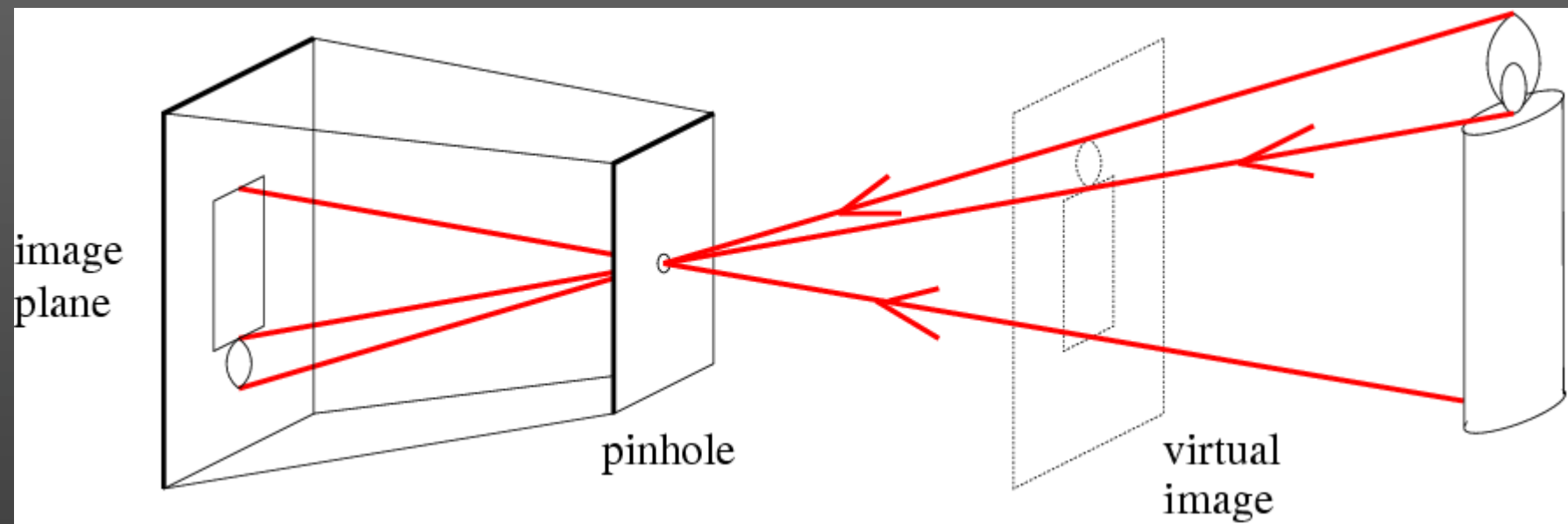
Outline

- What computer vision is about
- What this class is about
- A brief history of visual recognition
- A brief recap on geometry

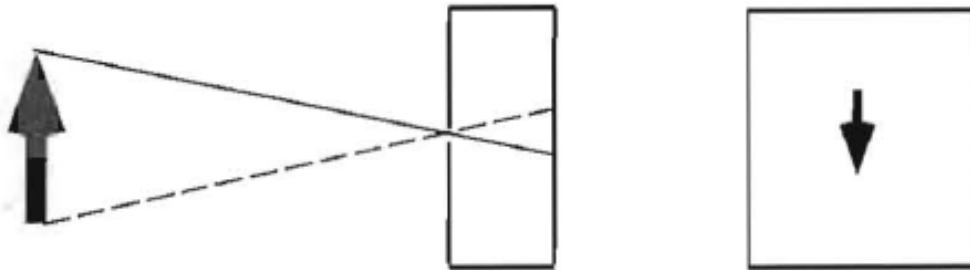
They are formed by the projection of three-dimensional objects.



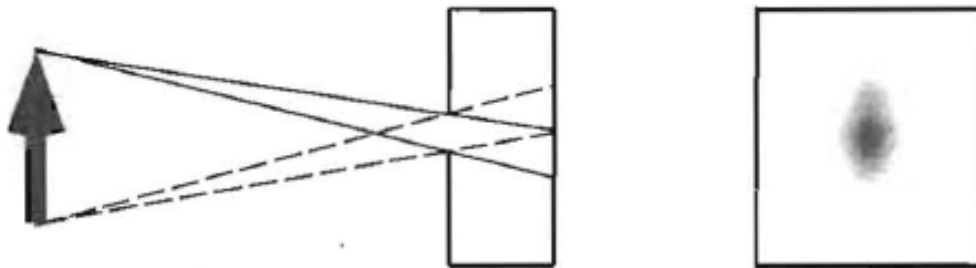
Images are brightness/color patterns drawn in a plane.



Pinhole camera: trade-off between sharpness and light transmission



A. Pinhole Aperture without Lens --> Sharp Image



B. Large Aperture without Lens --> Fuzzy Image

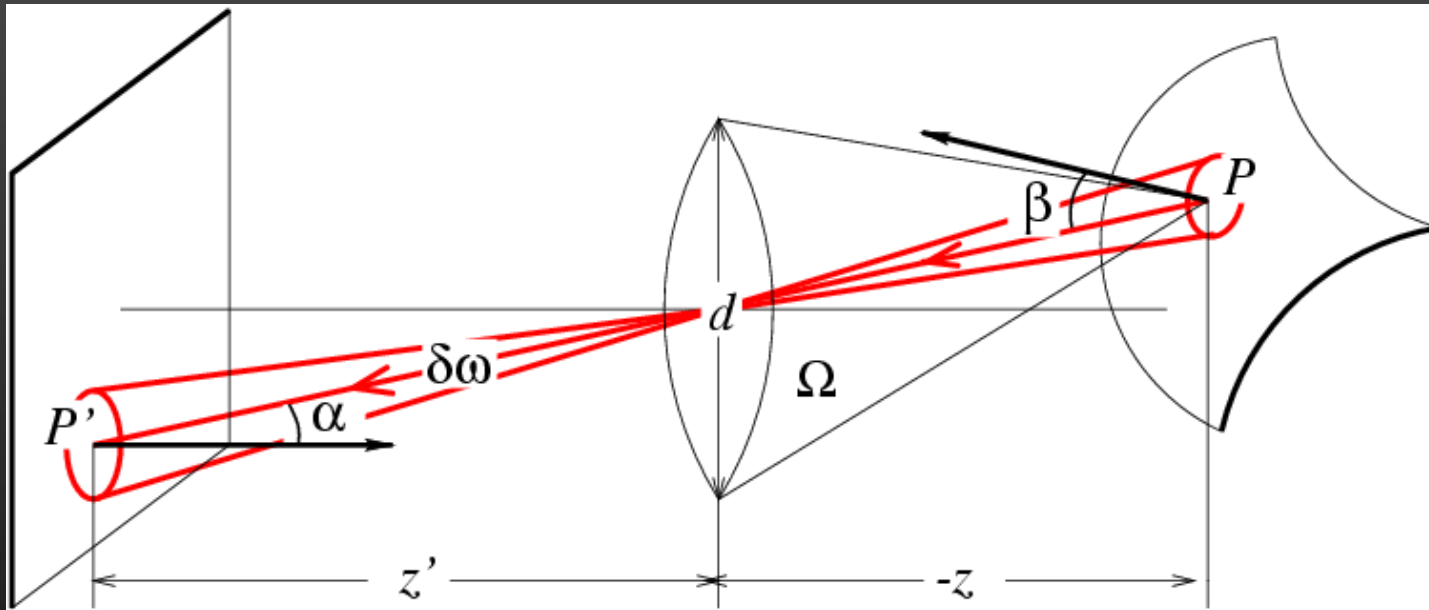


Camera Obscura in
Edinburgh

Advantages of lens systems

Lenses

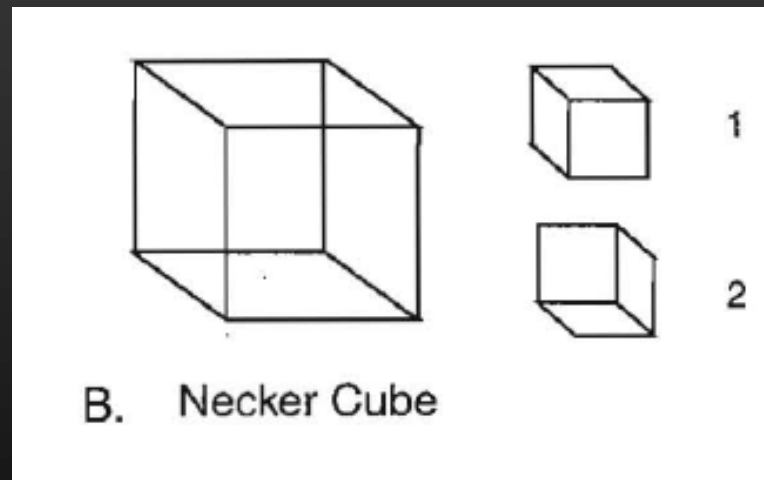
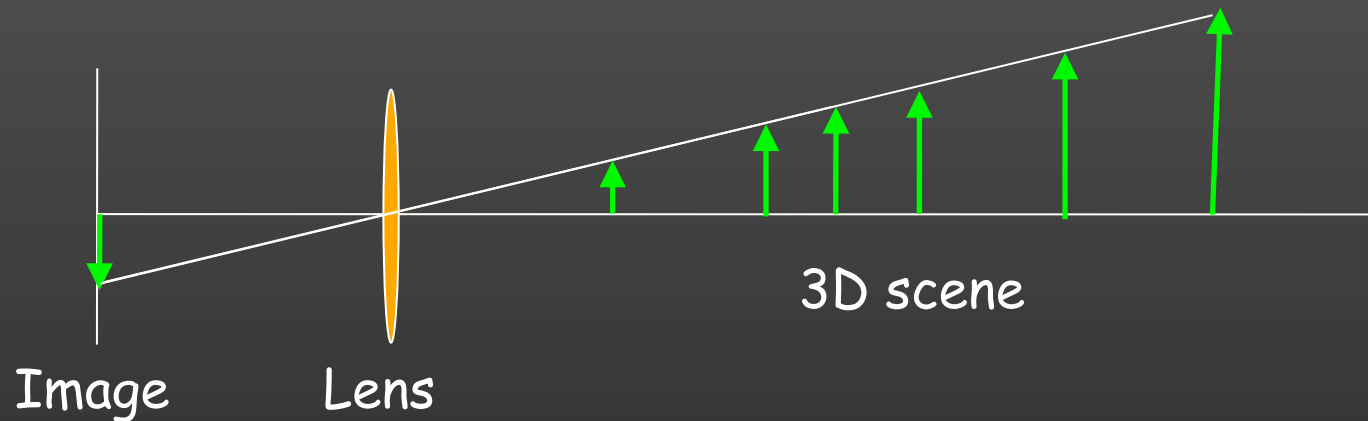
- can focus sharply on close and distant objects
- transmit more light than a pinhole camera



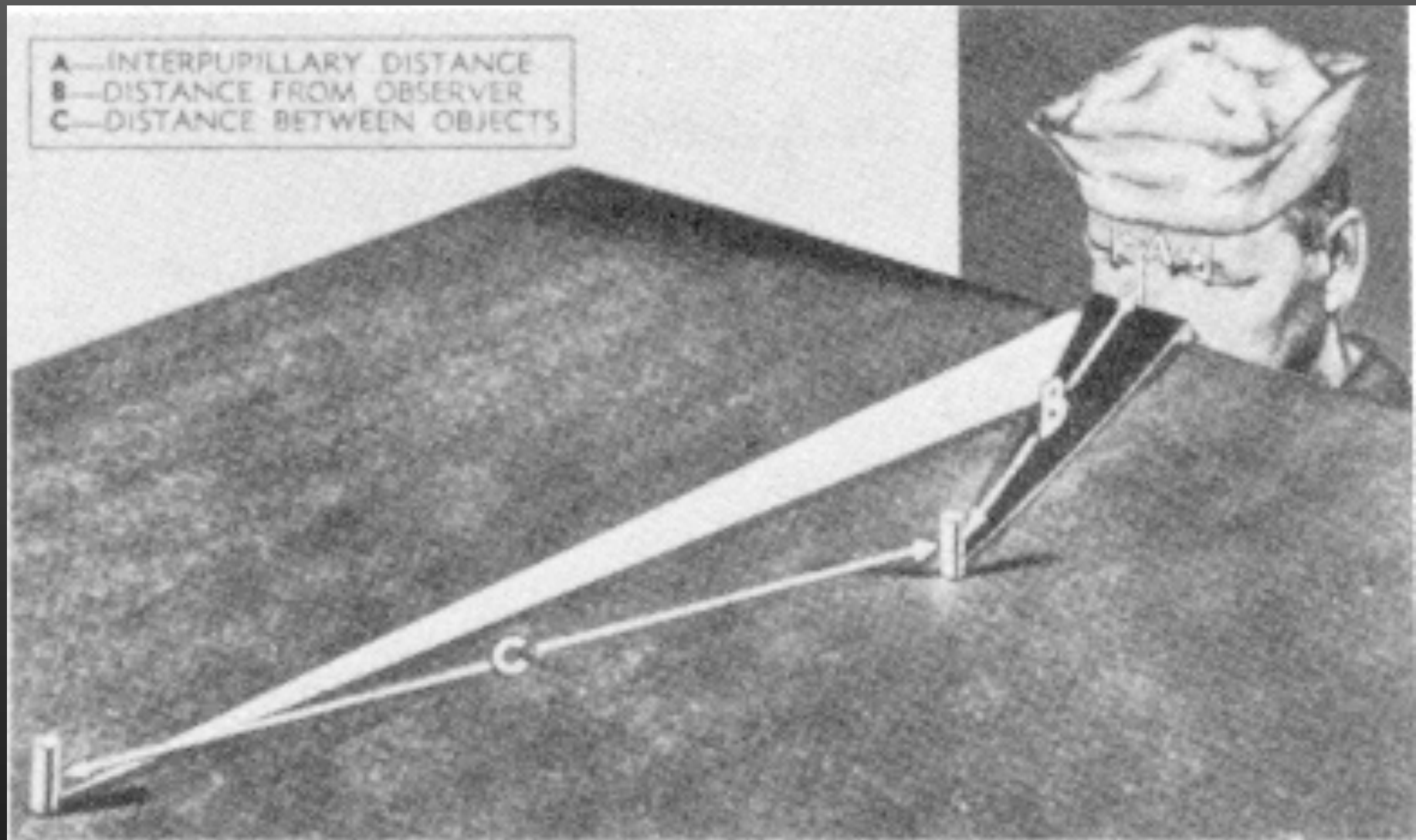
$$E = (\pi/4) \left[(d/z')^2 \cos^4 \alpha \right] L$$

Fundamental problem I: 3D world is "flattened" to 2D images

➔ Loss of information

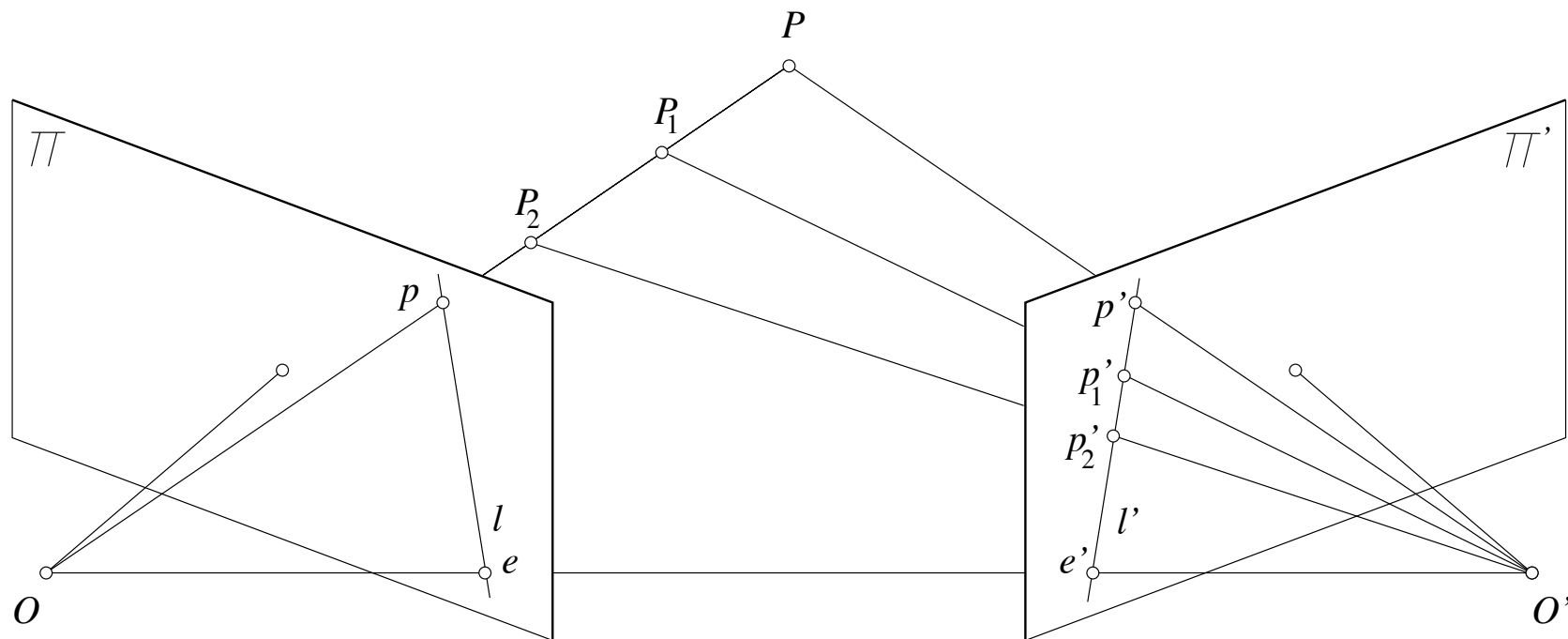


Question : how do we see "in 3D" ?

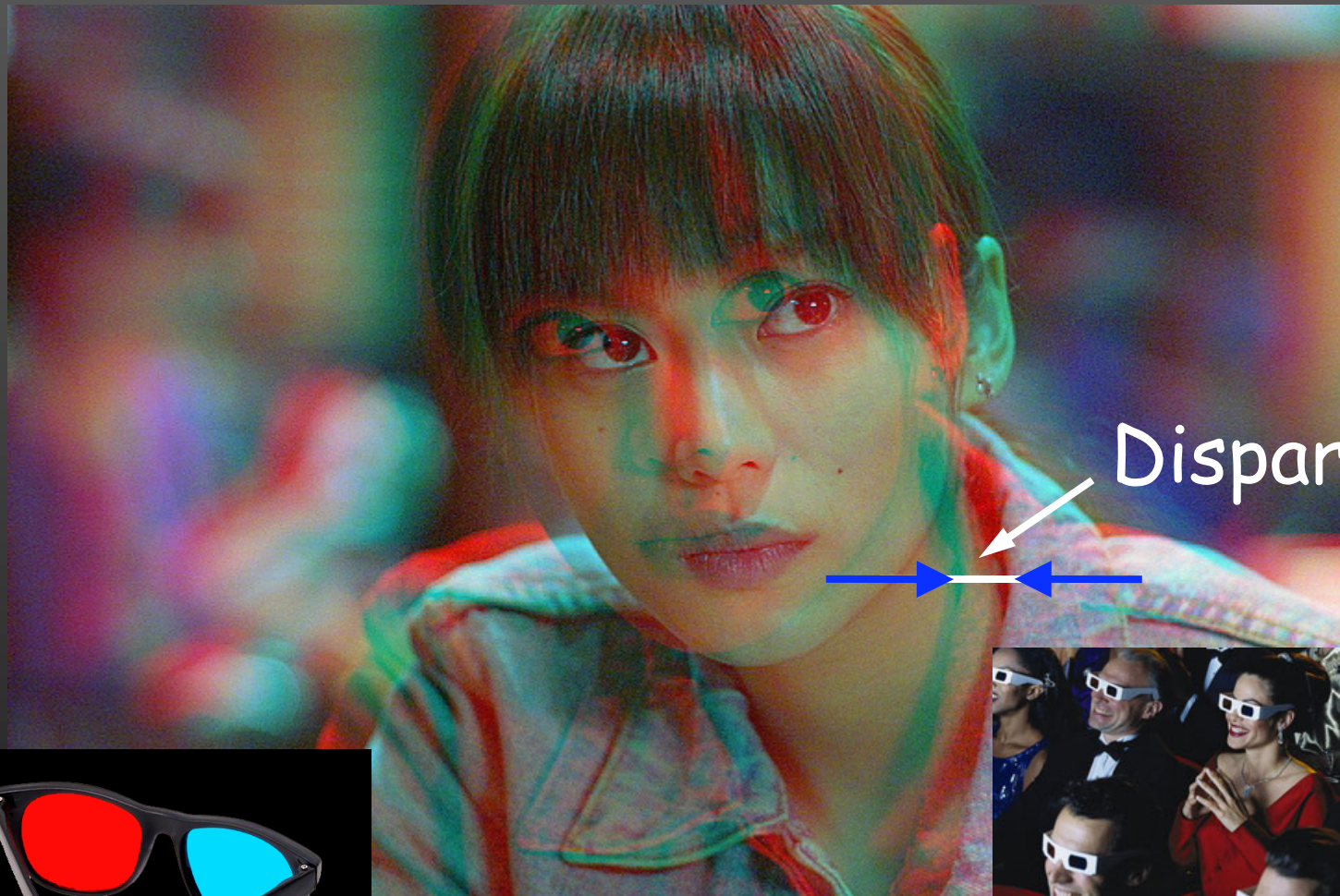


(First-order) answer: with our two eyes.

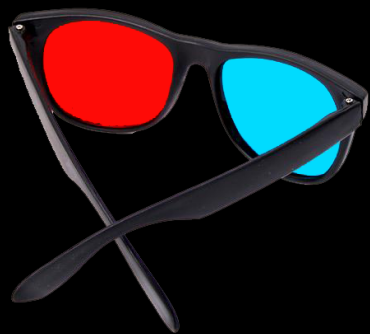
Epipolar Geometry



Simulated 3D perception

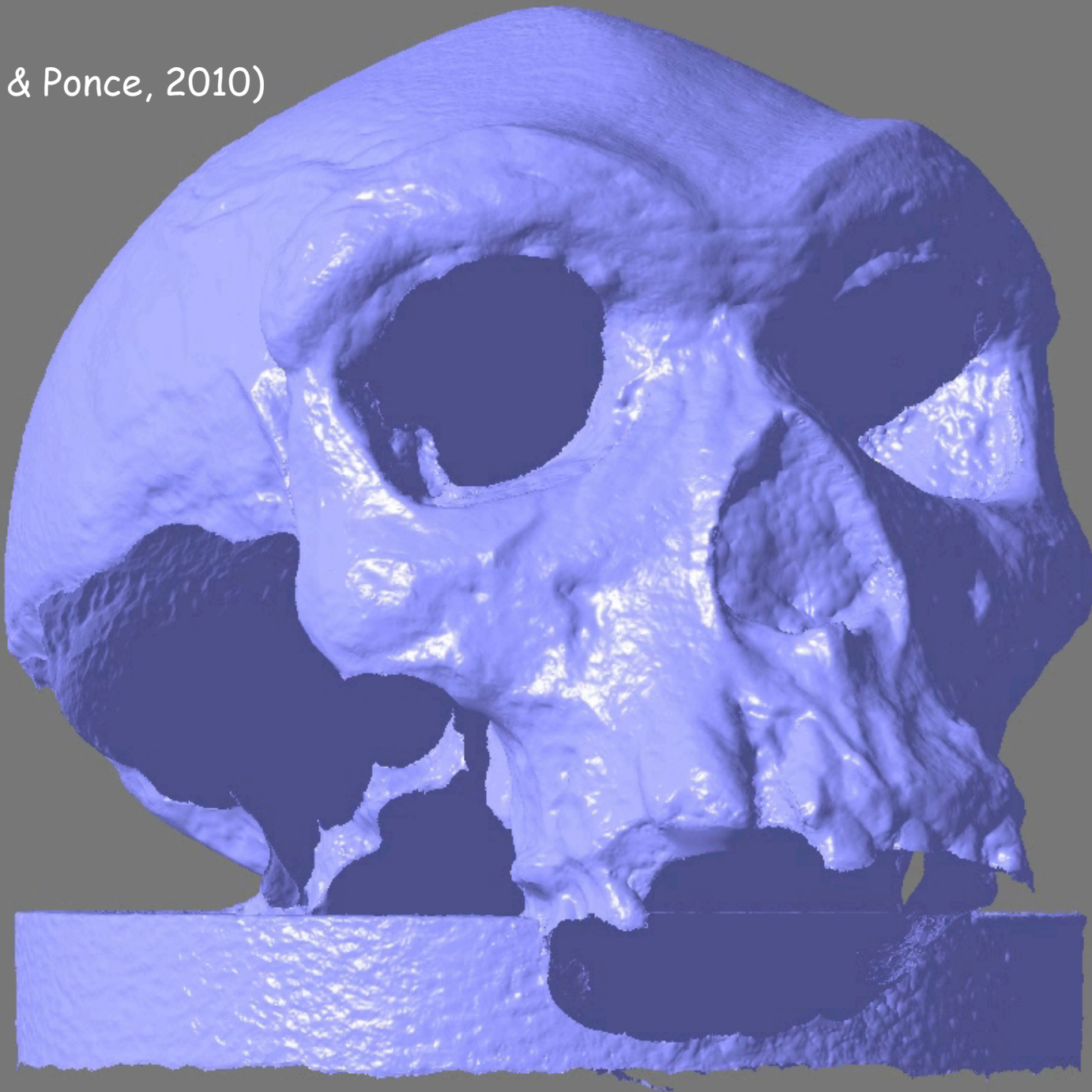


Disparity



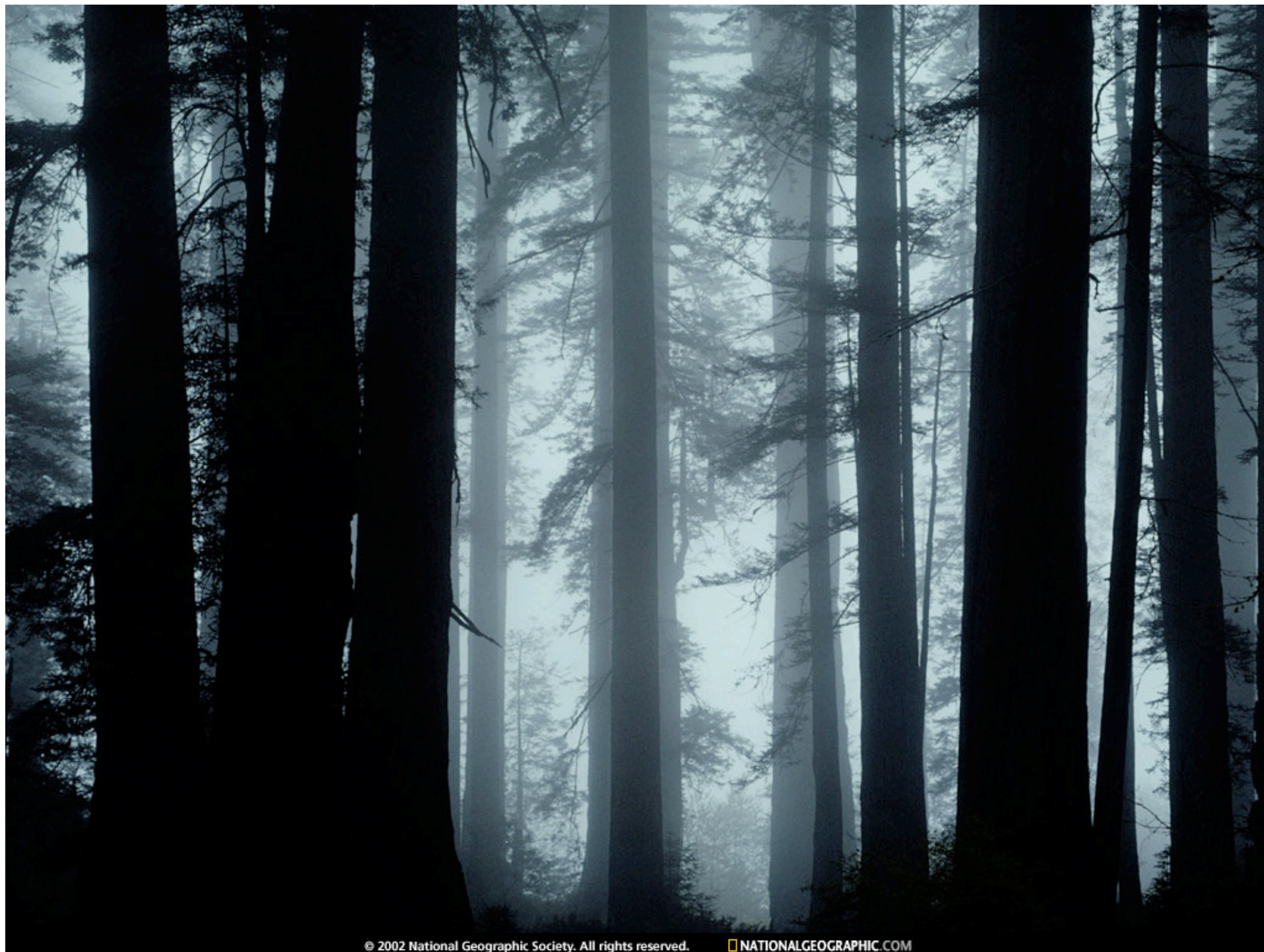
PMVS

(Furukawa & Ponce, 2010)



But there are other cues..





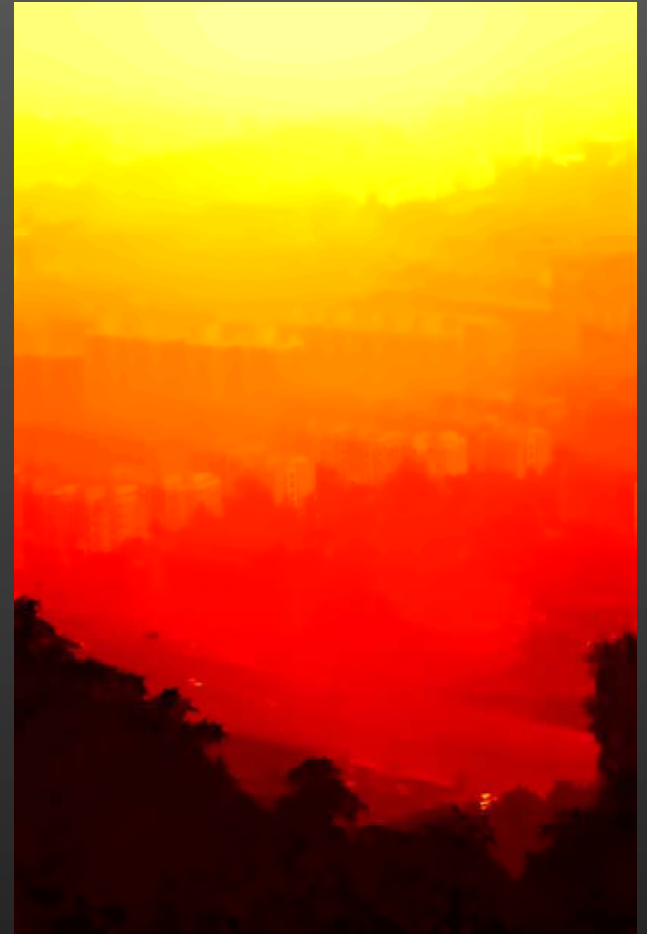
Depth from haze



Input haze image

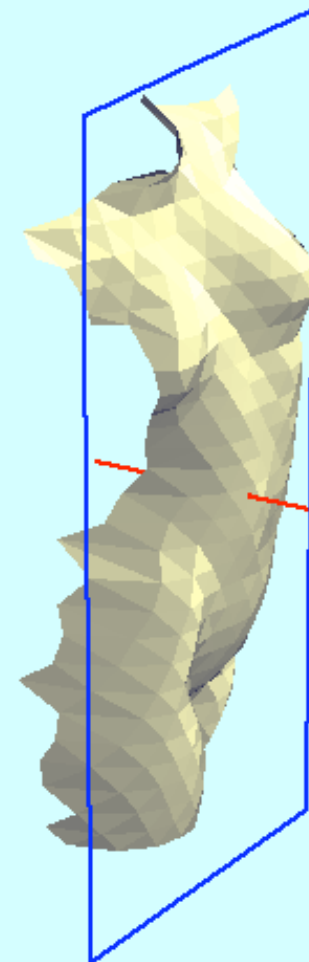


Reconstructed images



Recovered depth map

[K. HE, J. Sun and X. Tang, CVPR 2009]



Source: J. Koenderink



Source: J. Koenderink

What is happening with the shadows?





Image source: F. Durand

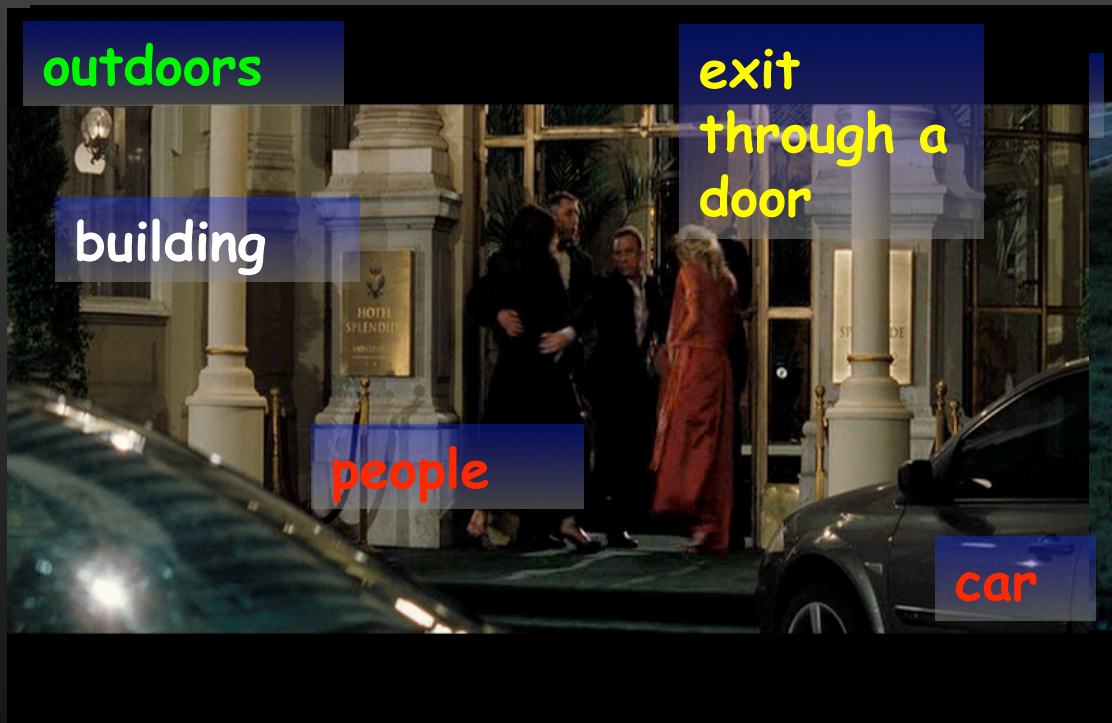
Challenges or opportunities?



Image source: J. Koenderink

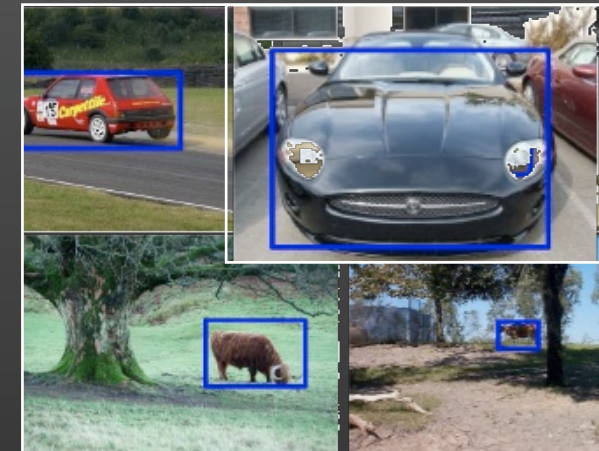
- Images are confusing, but they also reveal the structure of the world through numerous cues.
- Our job is to interpret the cues!

But we want much more than 3D:
ex: Visual scene analysis



How to make sense of "pixel-chaos"?

Object class recognition



3D Scene reconstruction



Face recognition



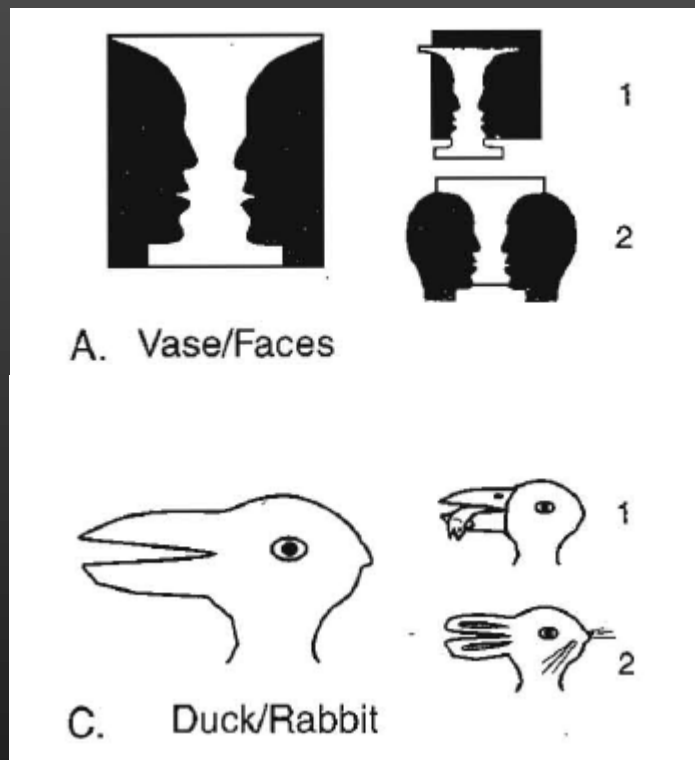
Action recognition



Drinking

Fundamental problem II: Images do not measure the meaning

→ We need lots of prior knowledge to make meaningful interpretations of an image



Outline

- What computer vision is about
- What this class is about
- A brief history of visual recognition
- A brief recap on geometry

Specific object detection



(Lowe, 2004)

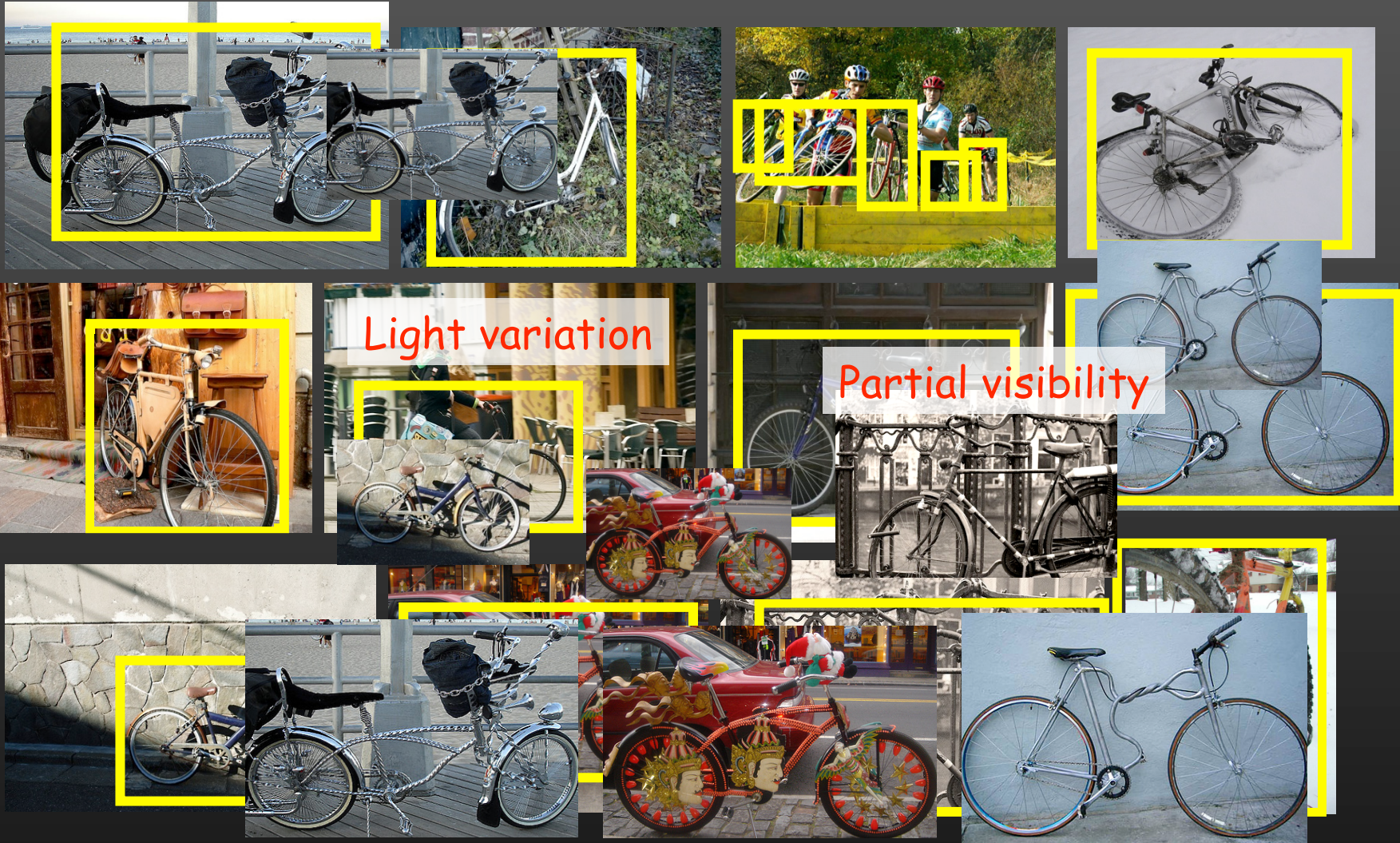
Image classification



Caltech 101 : http://www.vision.caltech.edu/Image_Datasets/Caltech101/

Object category detection

View variation



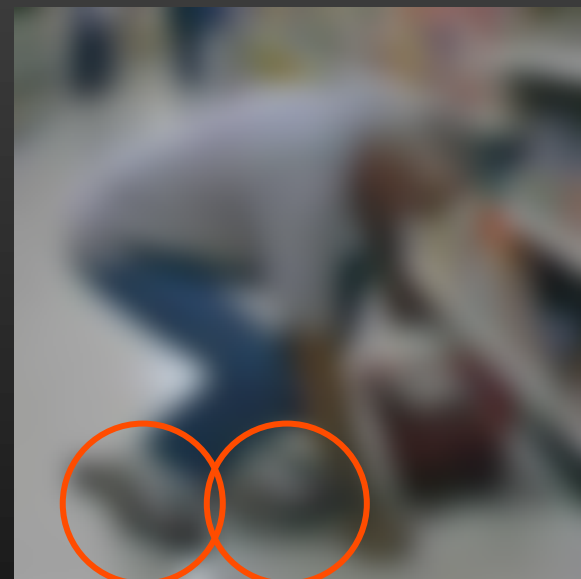
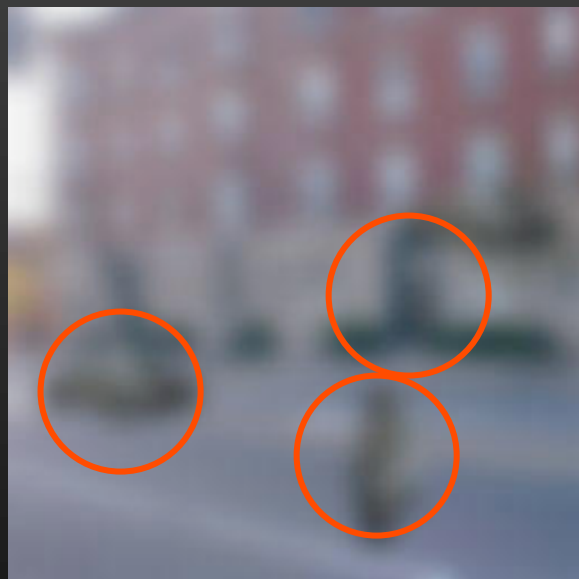
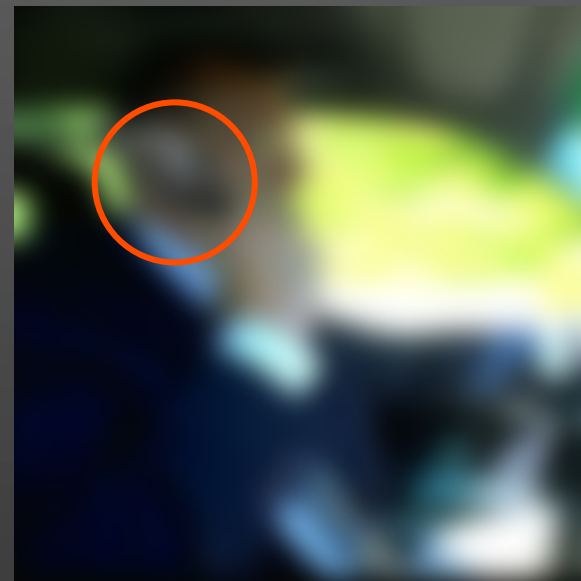
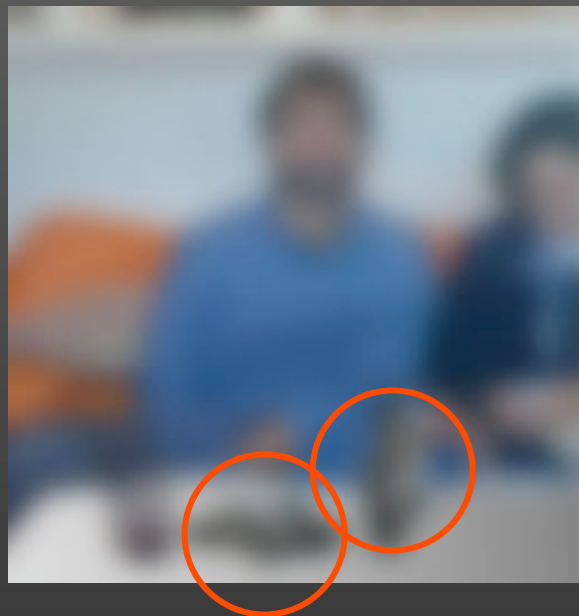
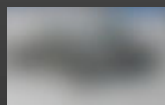
Within-class variation

Scene understanding



Photo courtesy A. Efros.

Local ambiguity and global scene interpretation



This class

1. Introduction plus recap on geometry (J. Ponce, J. Sivics)
2. Instance-level recognition I. - Local invariant features (C. Schmid)
3. Instance-level recognition II. - Correspondence, efficient visual search (J. Sivic)
4. Very large scale image indexing; bag-of-feature models for category-level recognition (C. Schmid)
5. Sparse coding (J. Ponce); object detection (J. Sivic)
6. Holiday, no lecture
7. Neural networks; optimization (N. Le Roux)
8. Object detection; pictorial structures; human pose (I. Laptev, J. Sivic)
9. Motion and human action (I. Laptev)
10. Face detection and recognition; segmentation (C. Schmid)
11. Scenes and objects (I. Laptev, J. Sivic)
12. Final project presentations (J. Sivic, I. Laptev)

Computer vision books

- D.A. Forsyth and J. Ponce, "Computer Vision: A Modern Approach, Prentice-Hall, 2003 (2nd edition coming up Oct. 2011).
- J. Ponce, M. Hebert, C. Schmid, and A. Zisserman, "Toward category-level object recognition", Springer LNCS, 2007.
- R. Szeliski, "Computer Vision: Algorithms and Applications", Springer, 2010.
- O. Faugeras, Q.T. Luong, and T. Papadopoulos, "Geometry of Multiple Images," MIT Press, 2001.
- R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision", Cambridge University Press, 2004.
- J. Koenderink, "Solid Shape", MIT Press, 1990.

Class web-page

<http://www.di.ens.fr/willow/teaching/recvis11>

Slides available after classes:

<http://www.di.ens.fr/willow/teaching/recvis11/lecture01.pptx>

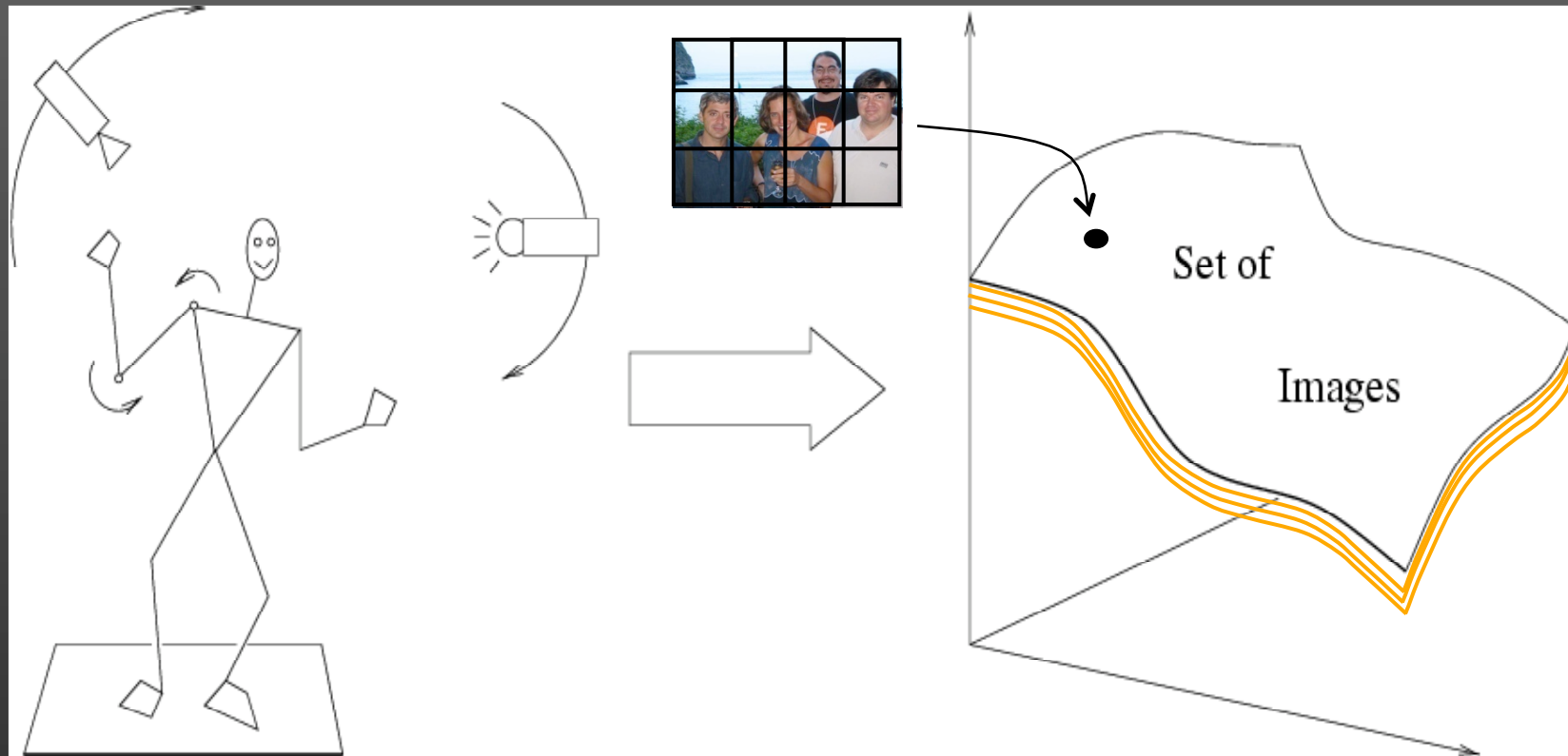
<http://www.di.ens.fr/willow/teaching/recvis11/lecture01.pdf>

Note: Much of the material used in this lecture is courtesy of Svetlana Lazebnik,

<http://www.cs.unc.edu/~lazebnik/>

Outline

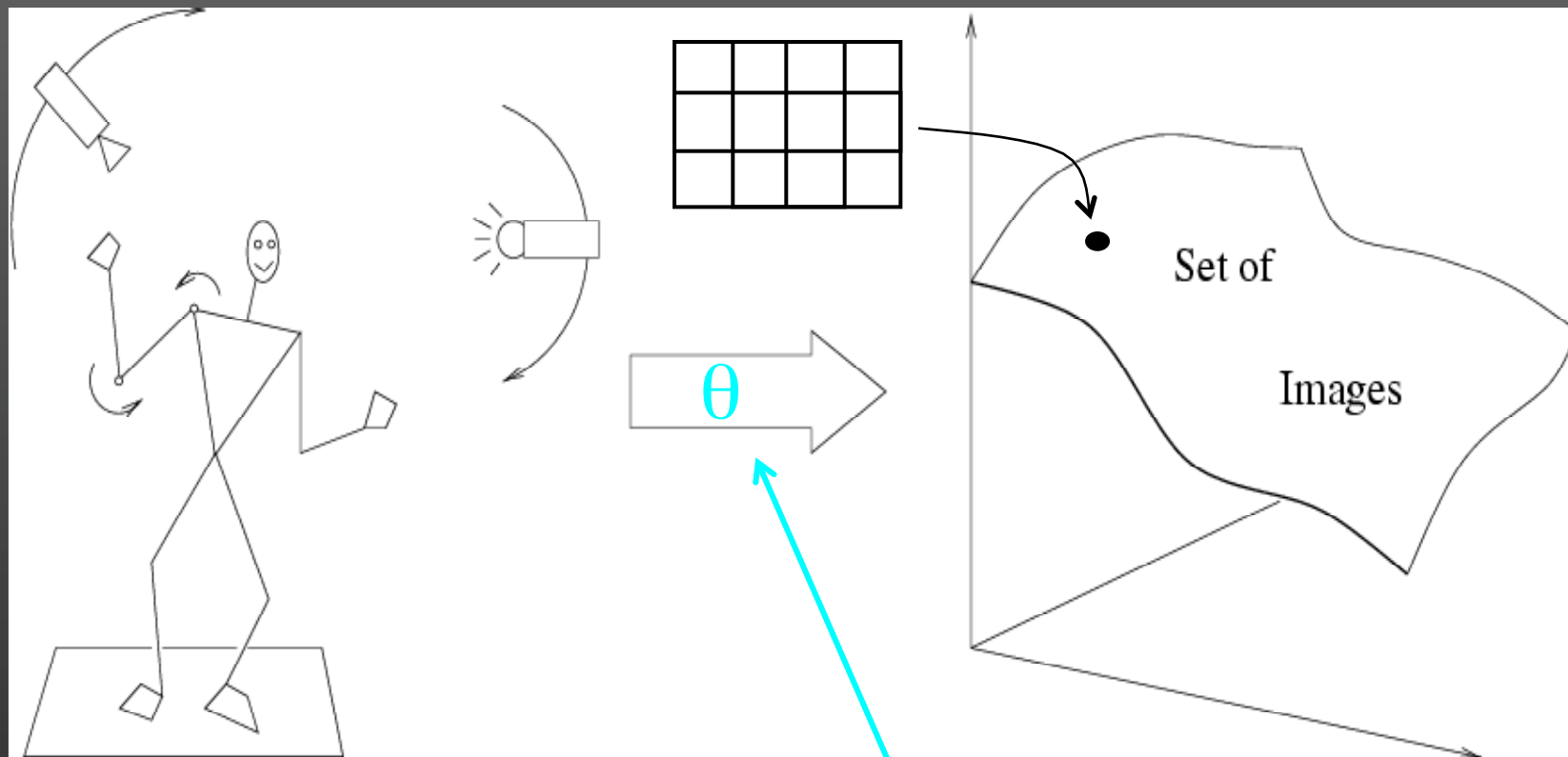
- What computer vision is about
- What this class is about
- A brief history of visual recognition
- A brief recap on geometry



Variability:



Camera position
Illumination
Internal parameters
Within-class variations

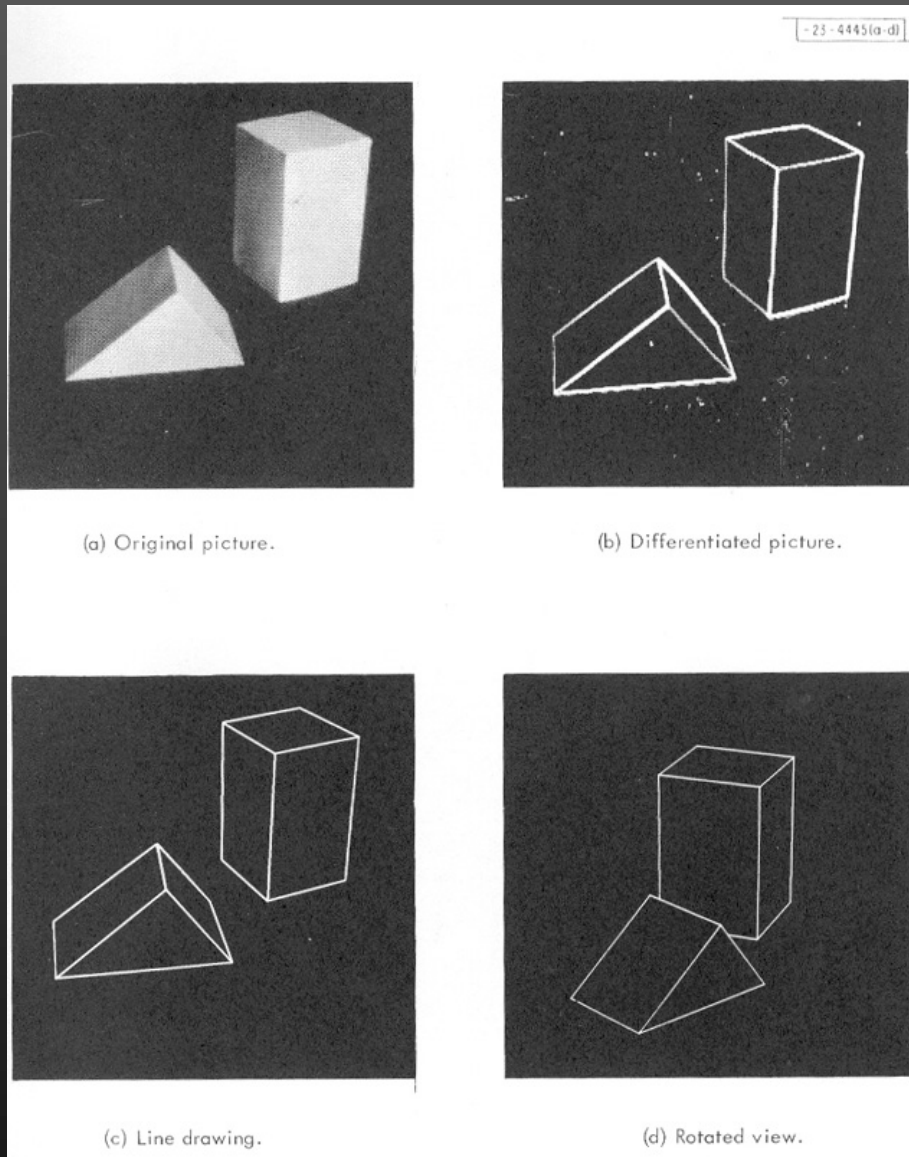


Variability:

Camera position
Illumination
Internal parameters

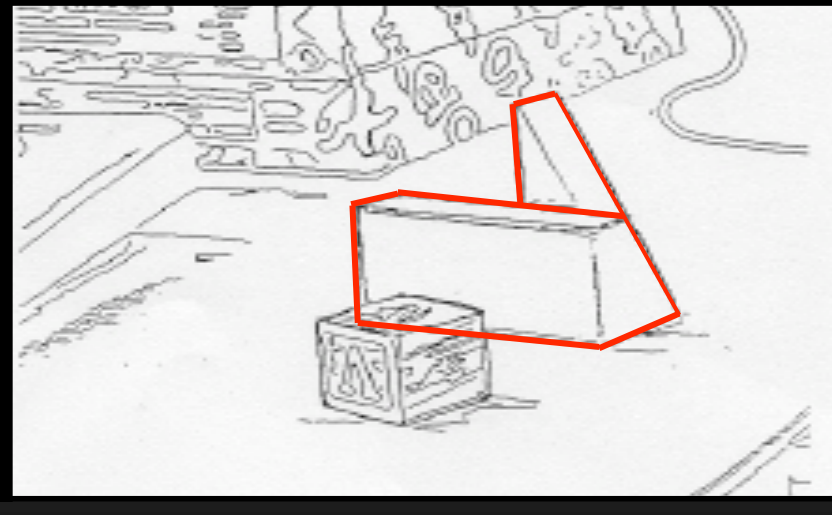
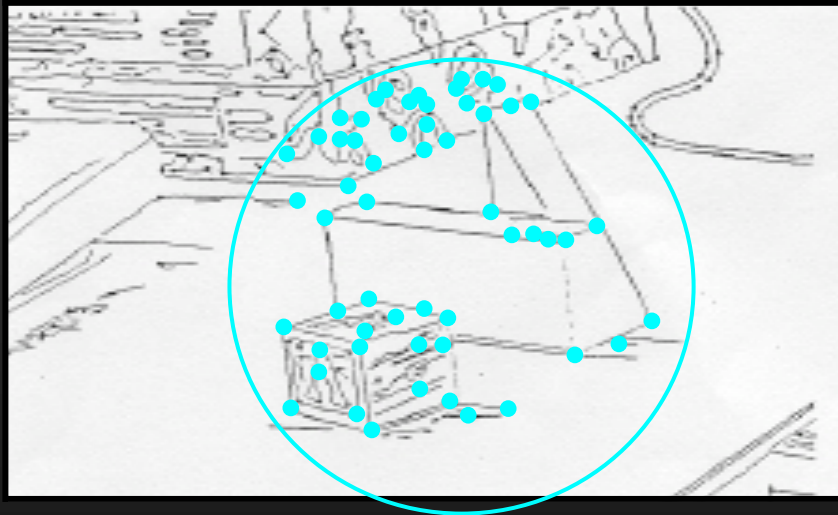
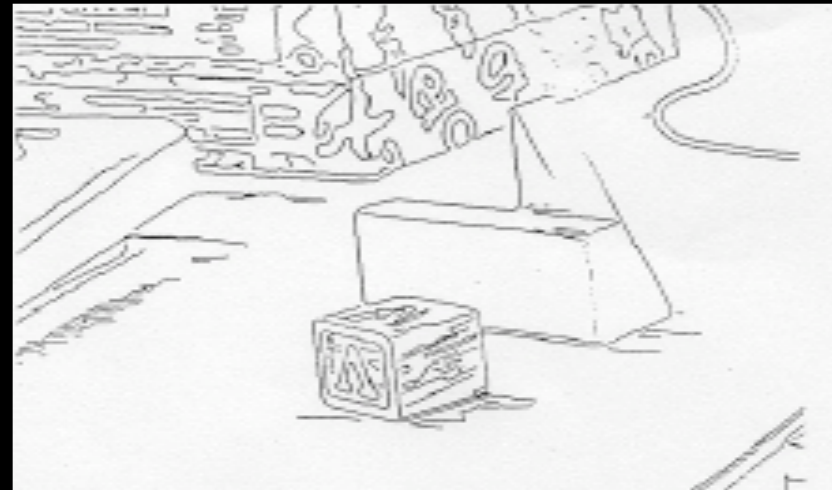
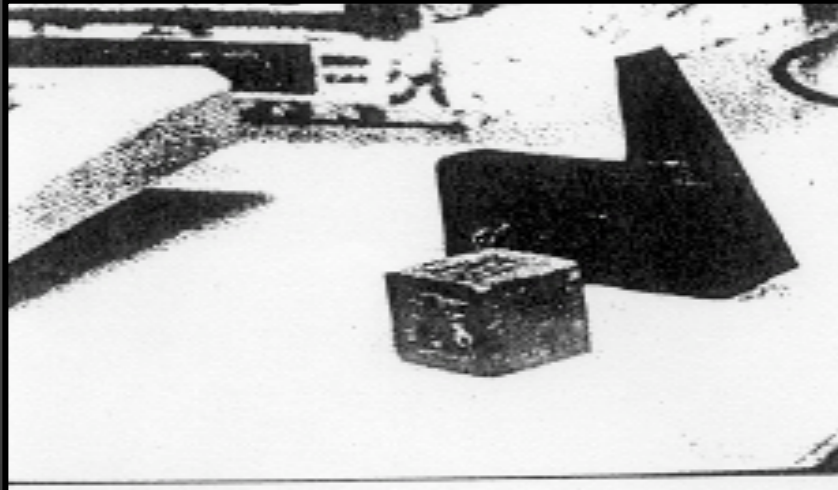
Roberts (1963); Lowe (1987); Faugeras & Hebert (1986); Grimson & Lozano-Perez (1986); Huttenlocher & Ullman (1987)

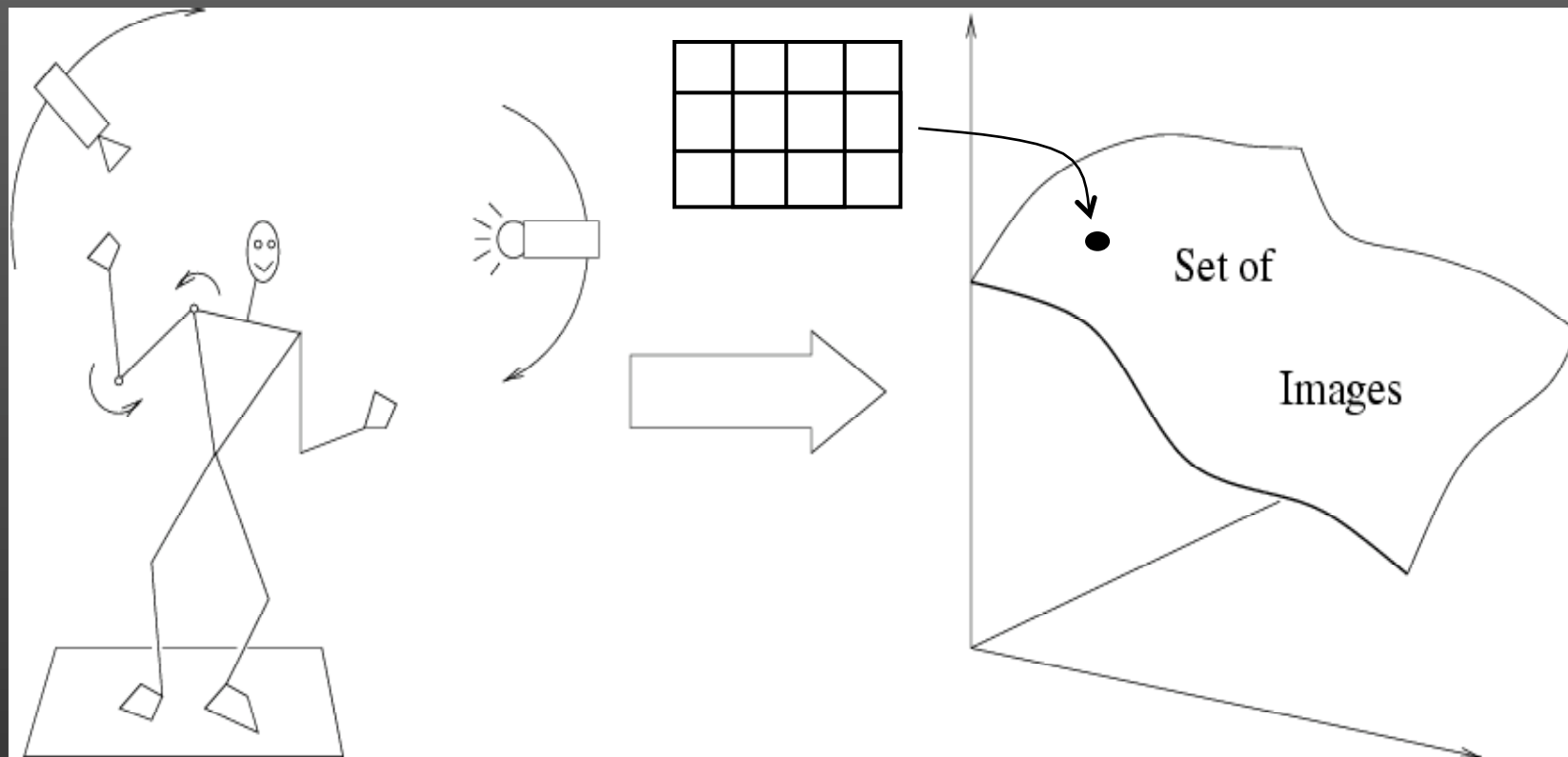
Origins of computer vision



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

Huttenlocher & Ullman (1987)





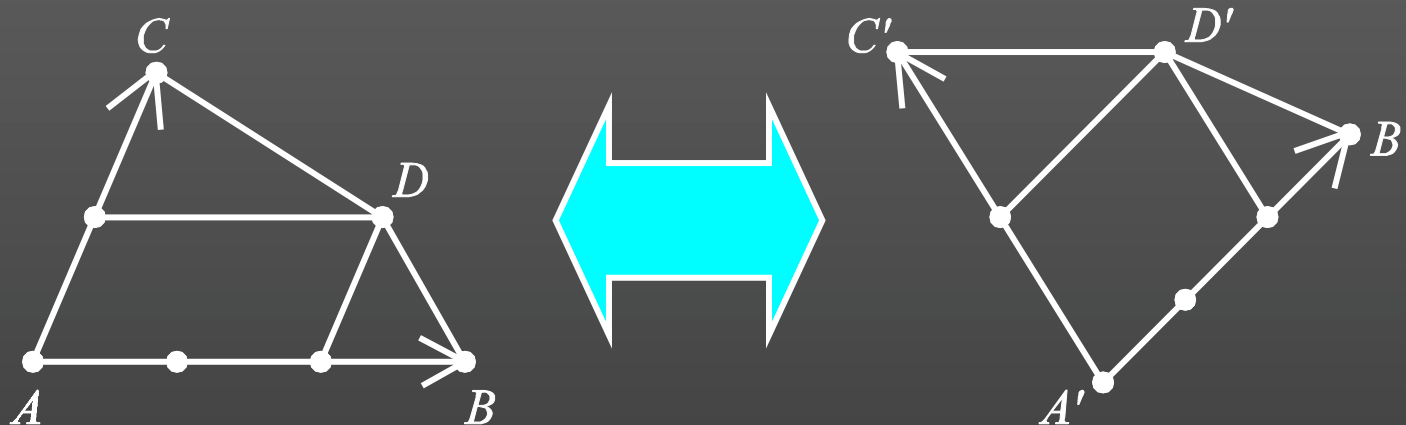
~~Variability~~

Invariance to:

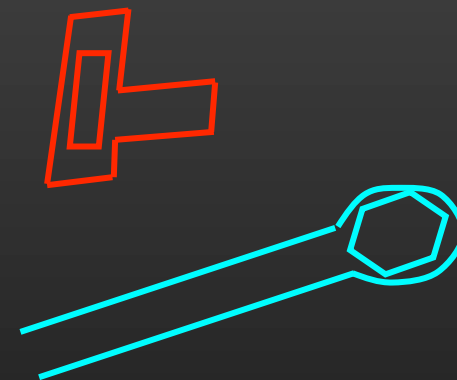
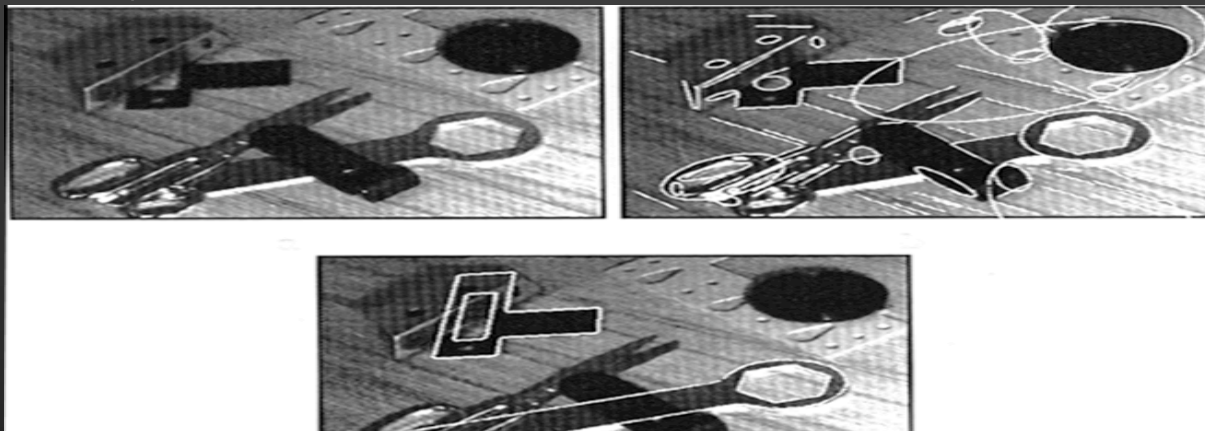
Camera position
Illumination
Internal parameters

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

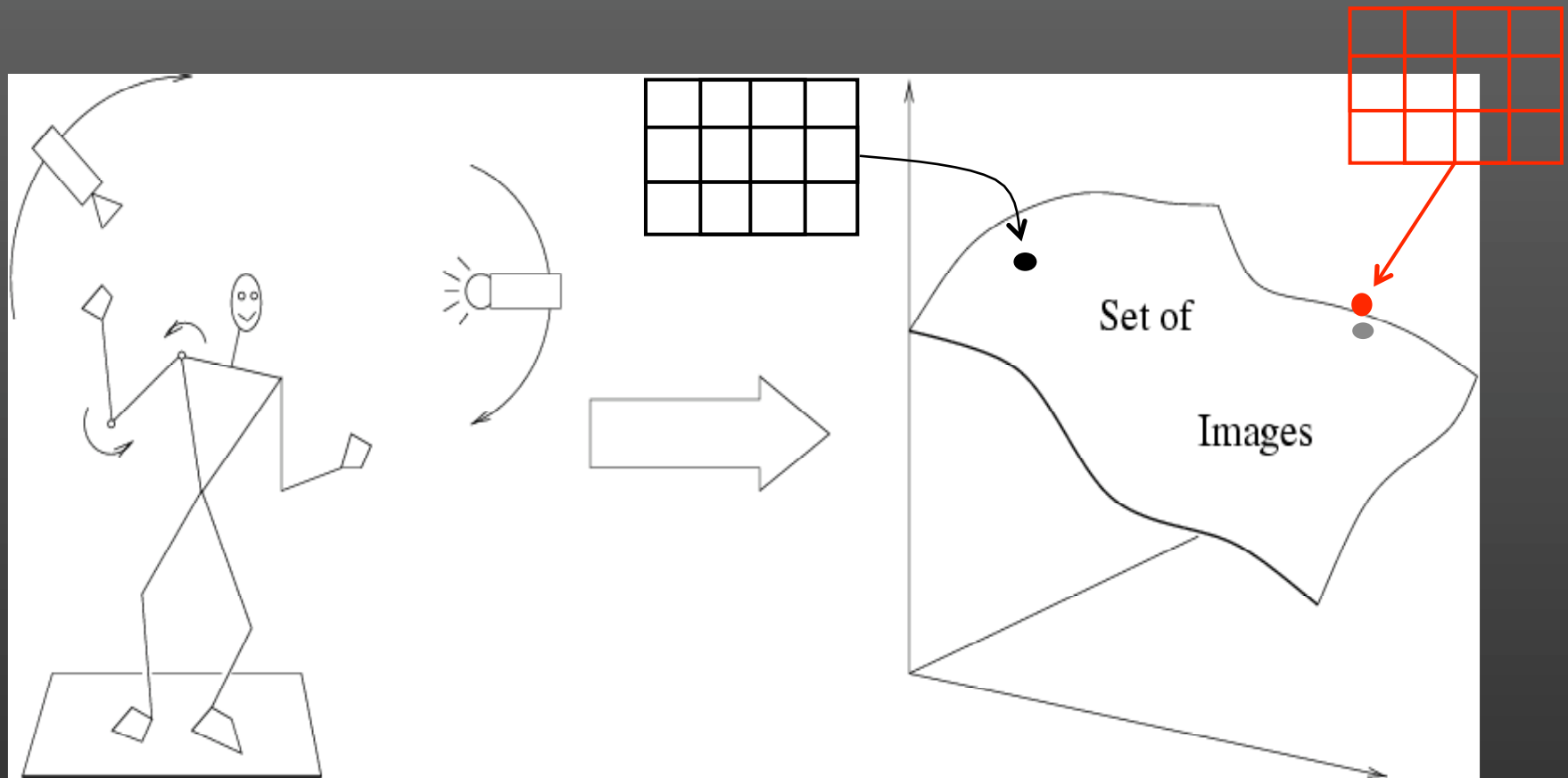
Example: affine invariants of coplanar points



Projective invariants (Rothwell et al., 1992):



BUT: True 3D objects do not admit monocular viewpoint invariants (Burns et al., 1993) !!



Empirical models of image variability:

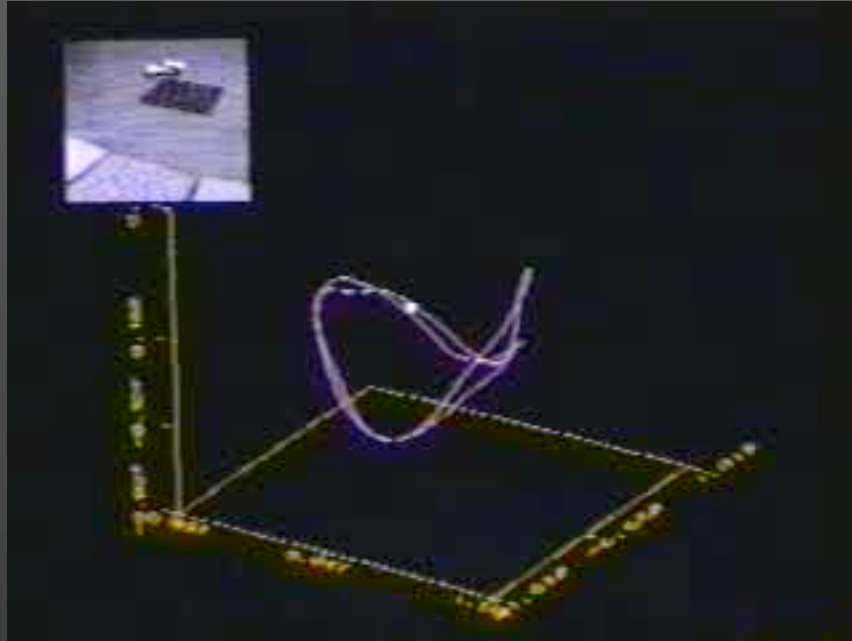
Appearance-based techniques

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

Eigenfaces (Turk & Pentland, 1991)



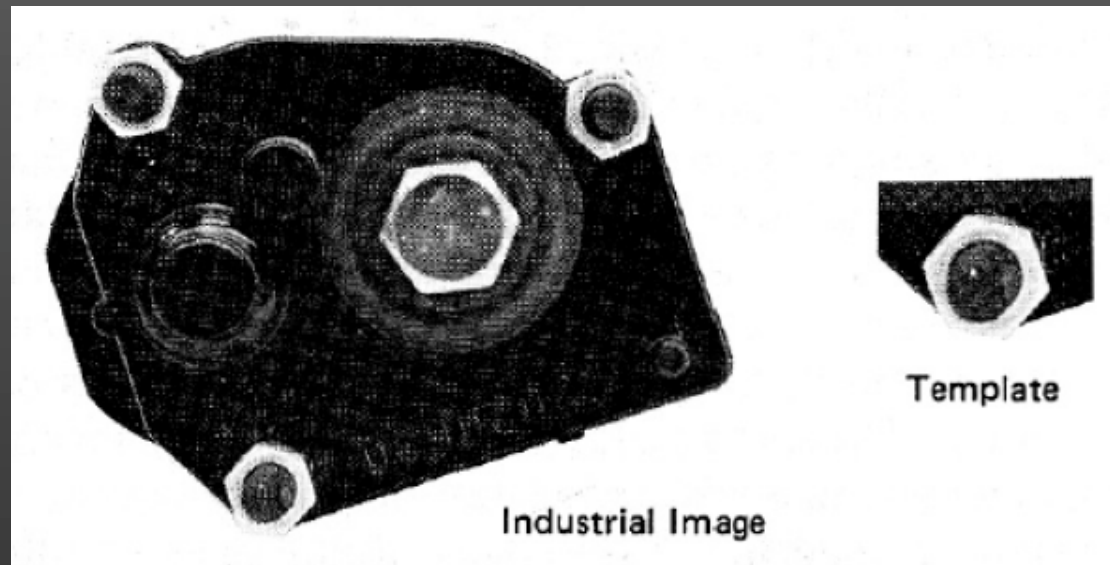
Experimental Condition	Correct/Unknown Recognition Percentage		
	Lighting	Orientation	Scale
Forced classification	96/0	85/0	64/0
Forced 100% accuracy	100/19	100/39	100/60
Forced 20% unknown rate	100/20	94/20	74/20



Appearance manifolds (Murase & Nayar, 1995)



Correlation-based template matching (60s)



Ballard & Brown (1980, Fig. 3.3). Courtesy Bob Fisher and Ballard & Brown on-line.

- Automated target recognition
- Industrial inspection
- Optical character recognition
- Stereo matching
- Pattern recognition

In the late 1990s, a new approach emerges:
Combining *local* appearance, spatial constraints, invariants,
and classification techniques from machine learning.

Query



Retrieved (10° off)

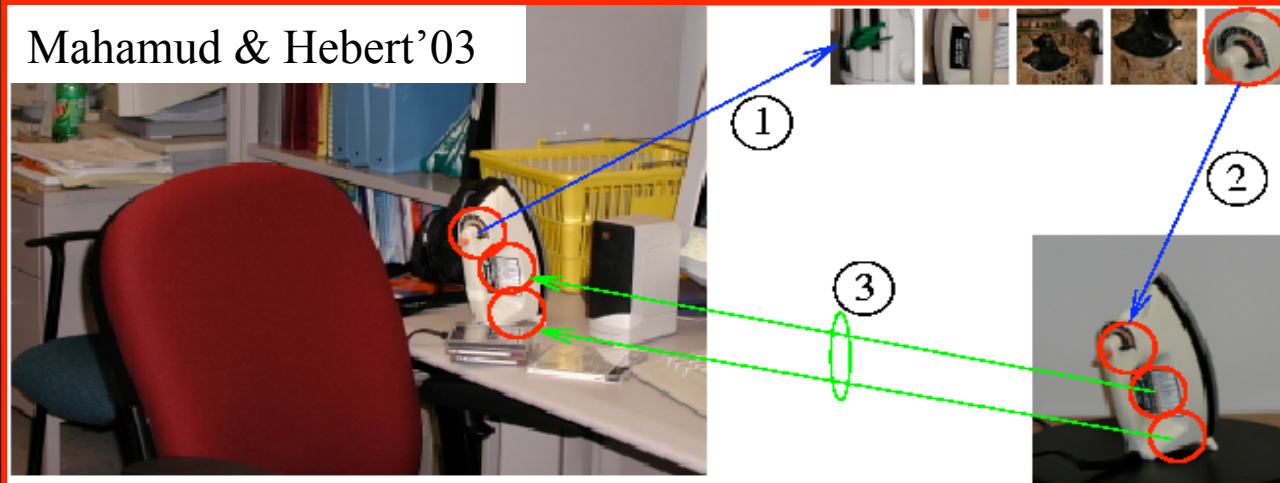


Schmid & Mohr'97

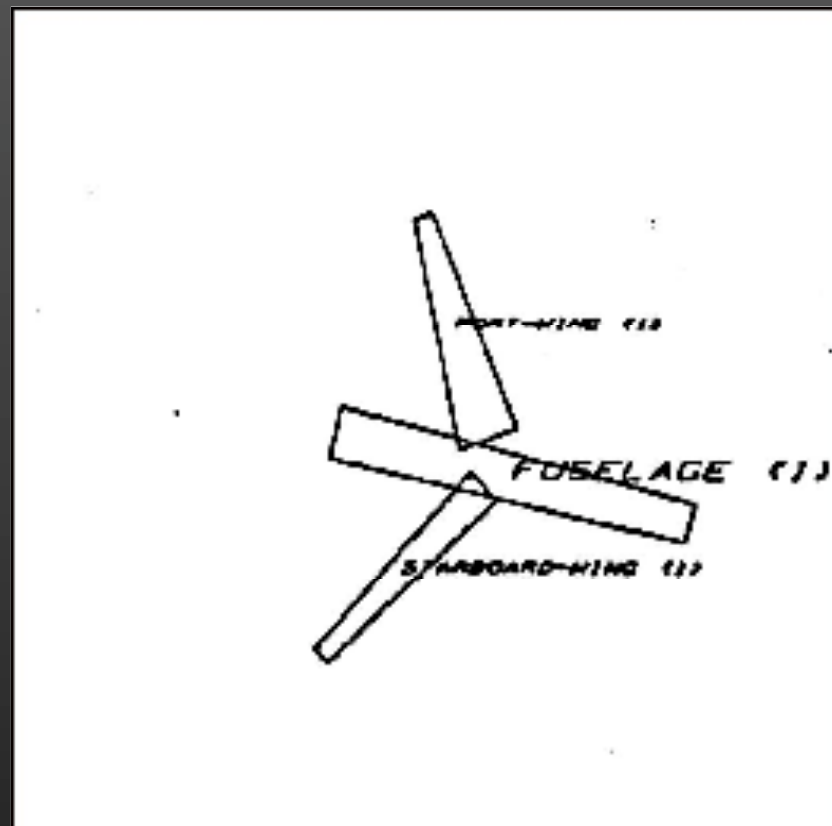
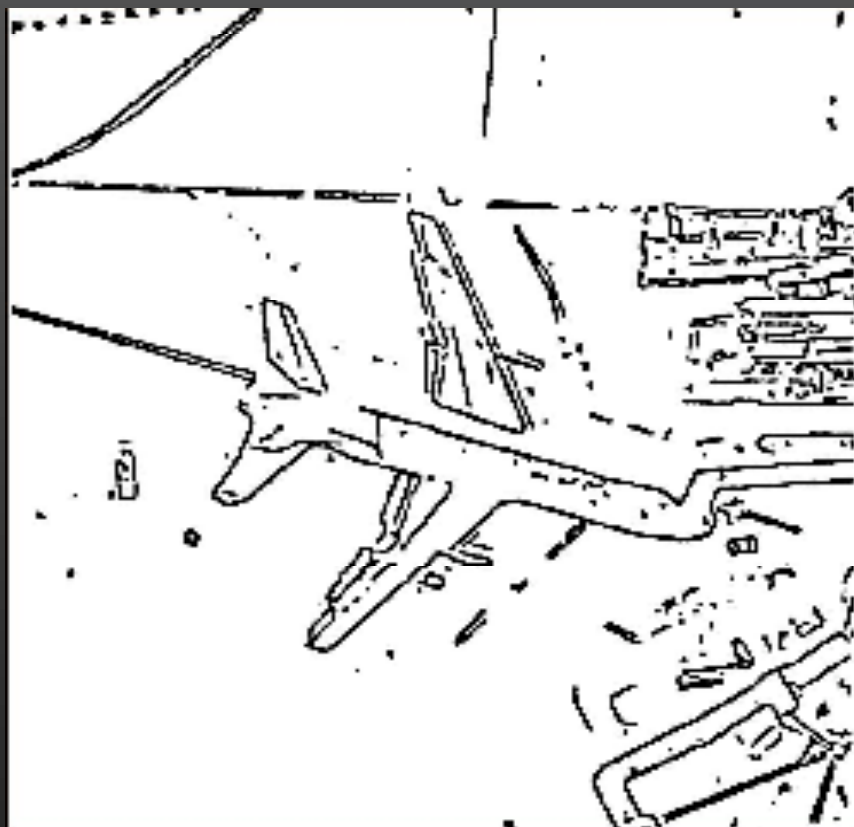
Lowe'02



Mahamud & Hebert'03



Representing and recognizing object categories is harder

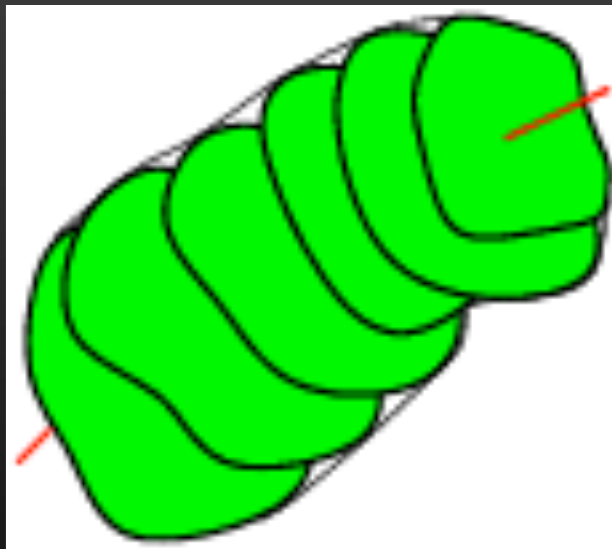
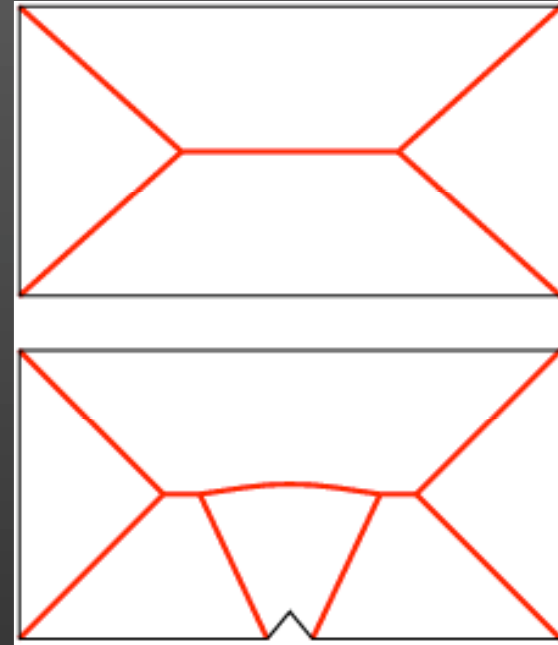
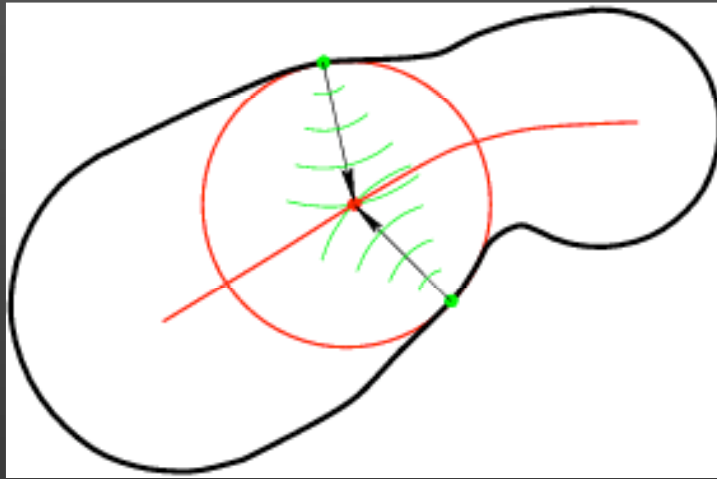


ACRONYM (Brooks and Binford, 1981)

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

Parts and invariants

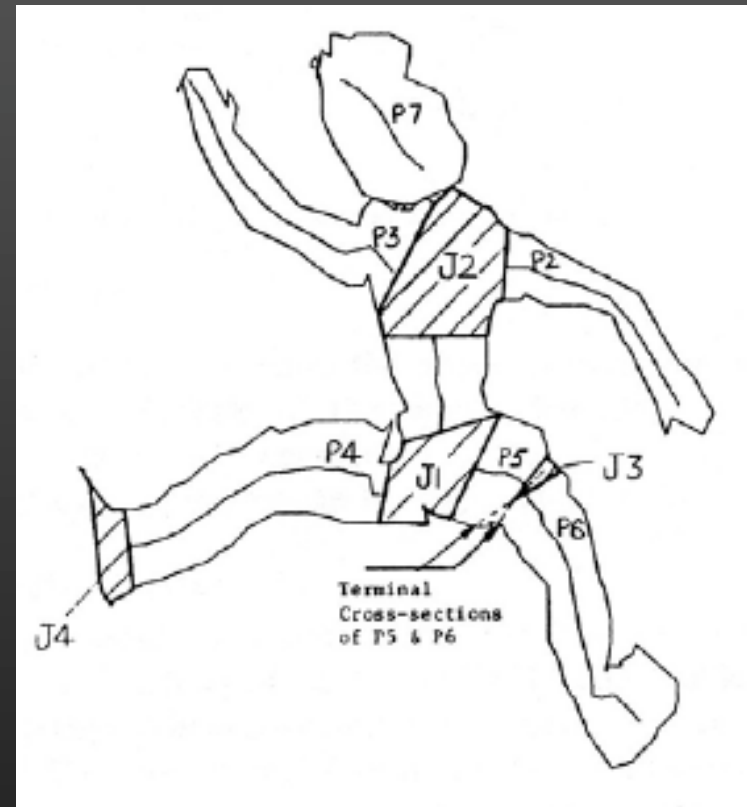
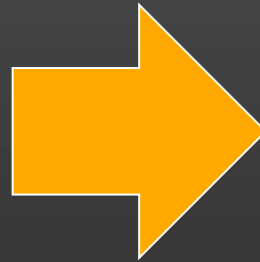
The Blum transform, 1967



Generalized cylinders
(Binford, 1971)

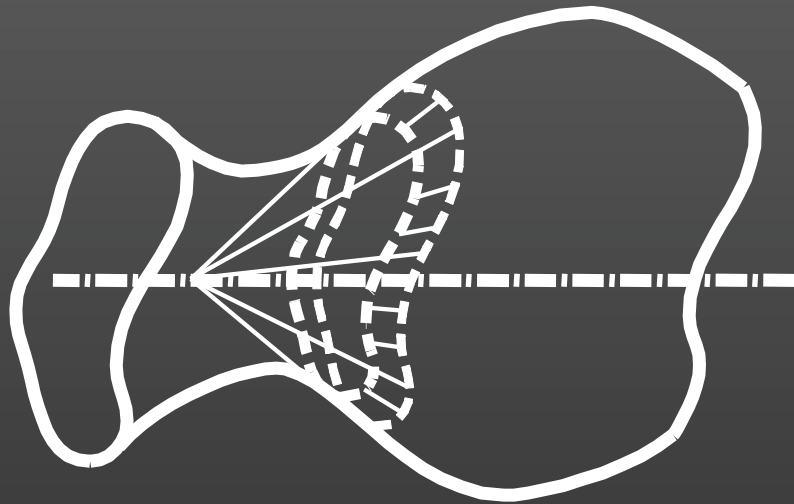
Generalized cylinders

(Binford, 1971; Marr & Nishihara, 1978)

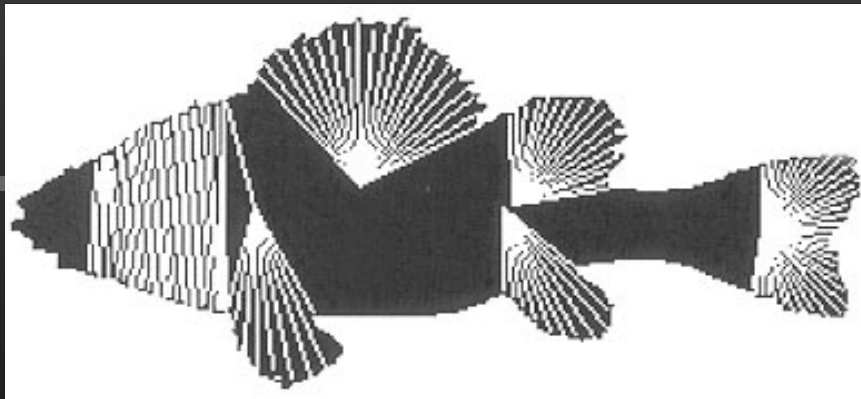


(Nevatia & Binford, 1972)

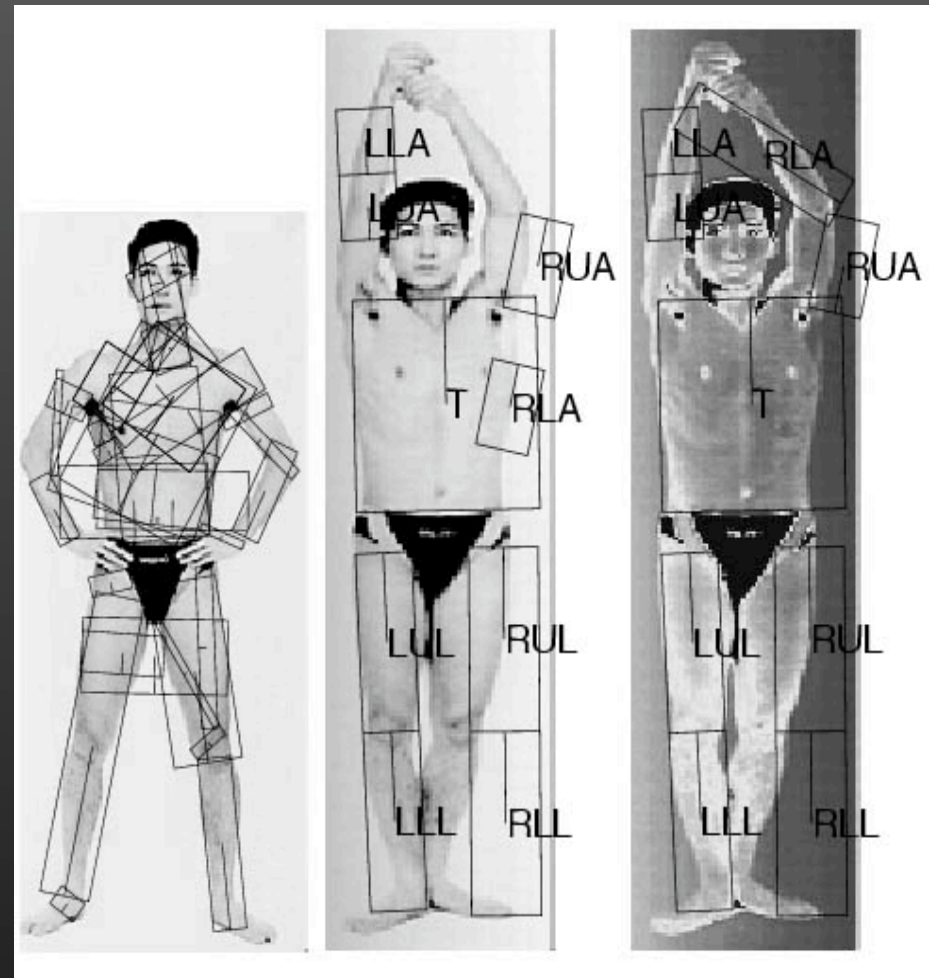
Parts and invariants II



Ponce et al. (1989)

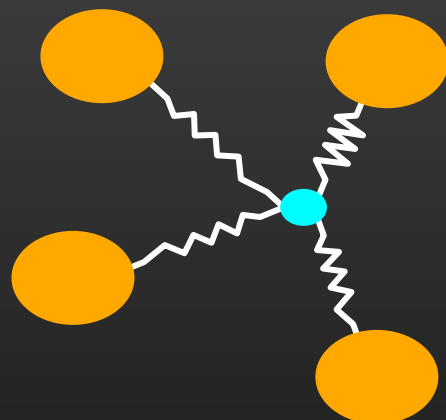
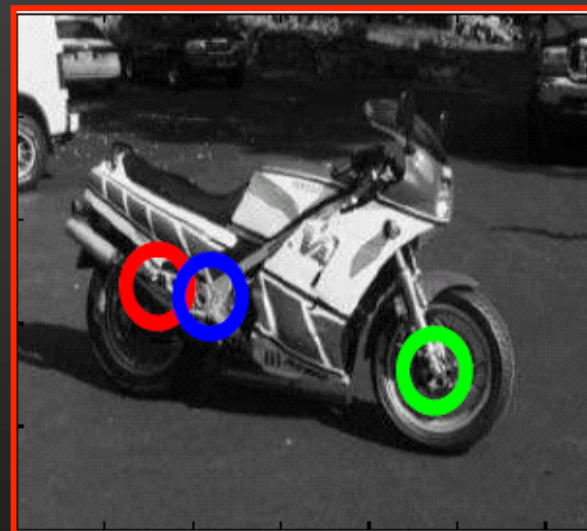
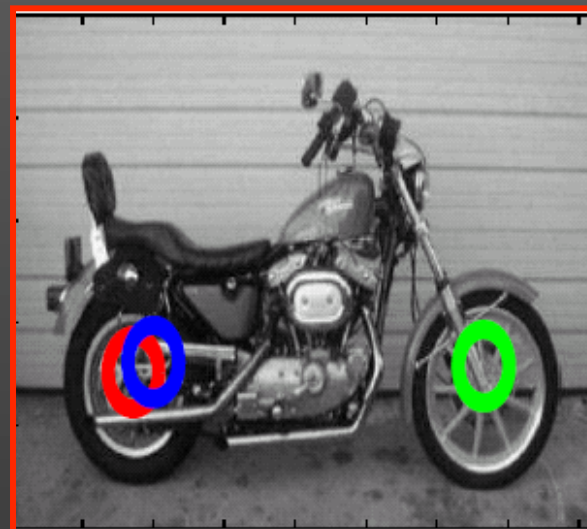


Zhu and Yuille (1996)



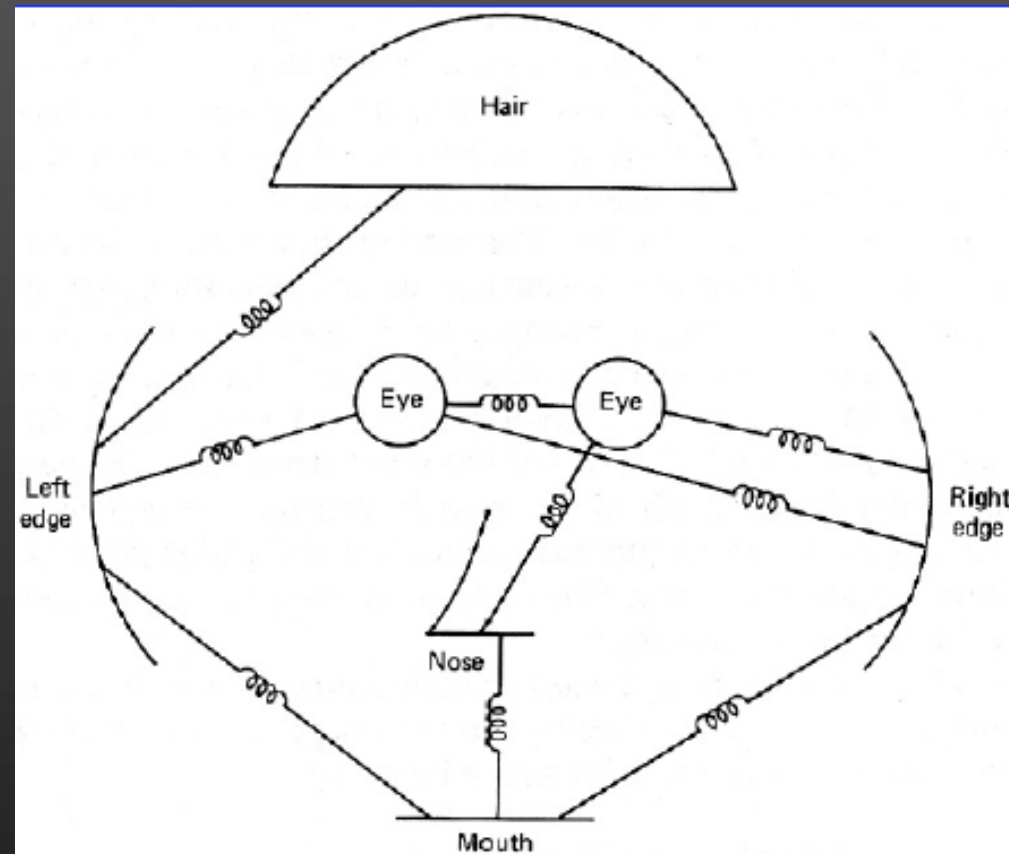
Ioffe and Forsyth (2000)

In the early 2000's, a new approach ?



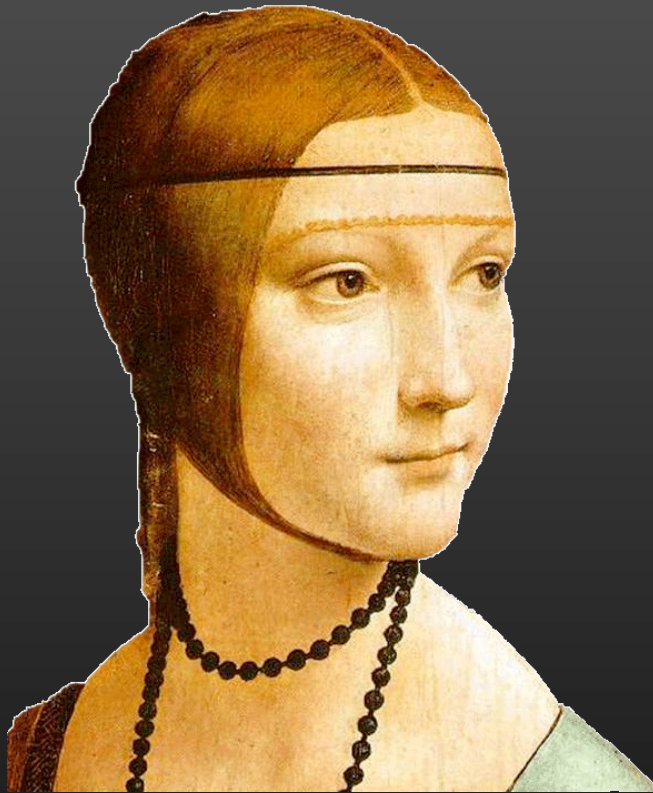
Fergus, Perona & Zisserman (2003)

The "templates and springs" model (Fischler & Elschlager, 1973)



Ballard & Brown (1980, Fig. 11.5). Courtesy
Bob Fisher and Ballard & Brown on-line.

Object $\longrightarrow$ Bag of 'words'



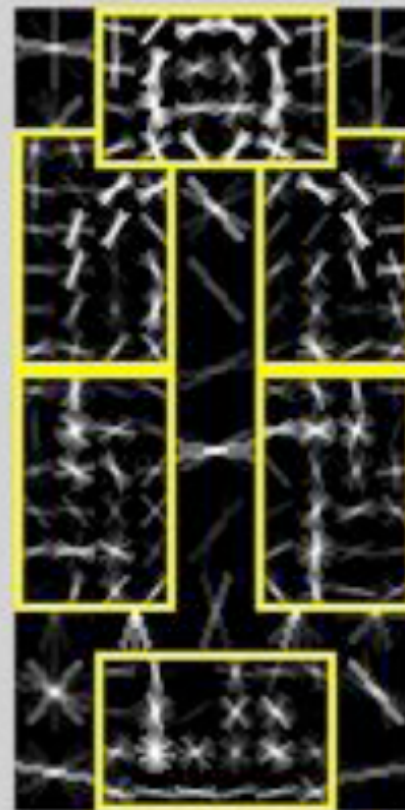
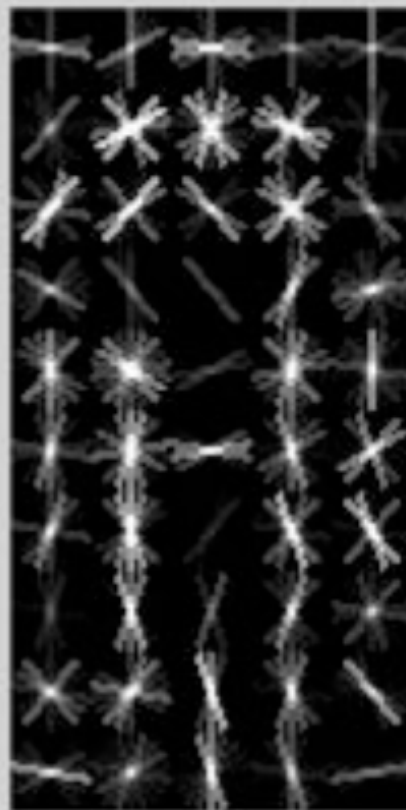


Color histograms (S&B'91)
Local jets (Florack'93)
Spin images (J&H'99)
Sift (Lowe'99)
Shape contexts (B&M'95)

Texon histograms (L&M'97)
Gist (O&T'05)
Spatial pyramids (LSP'06)
Hog (D&T'06)
Phog (B&Z'07)
Convolutional nets (LC'90)

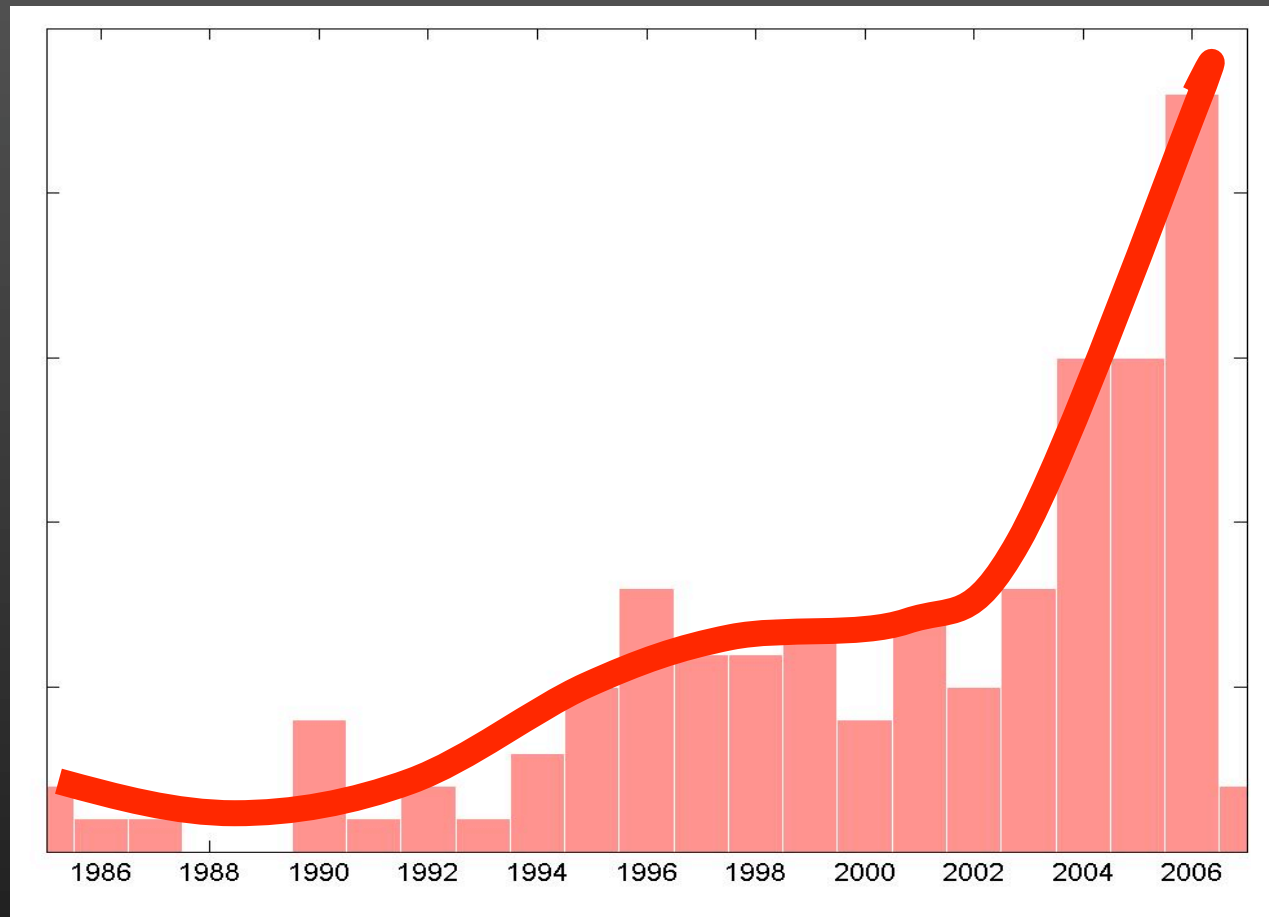


Locally orderless structure of images (K&vD'99)

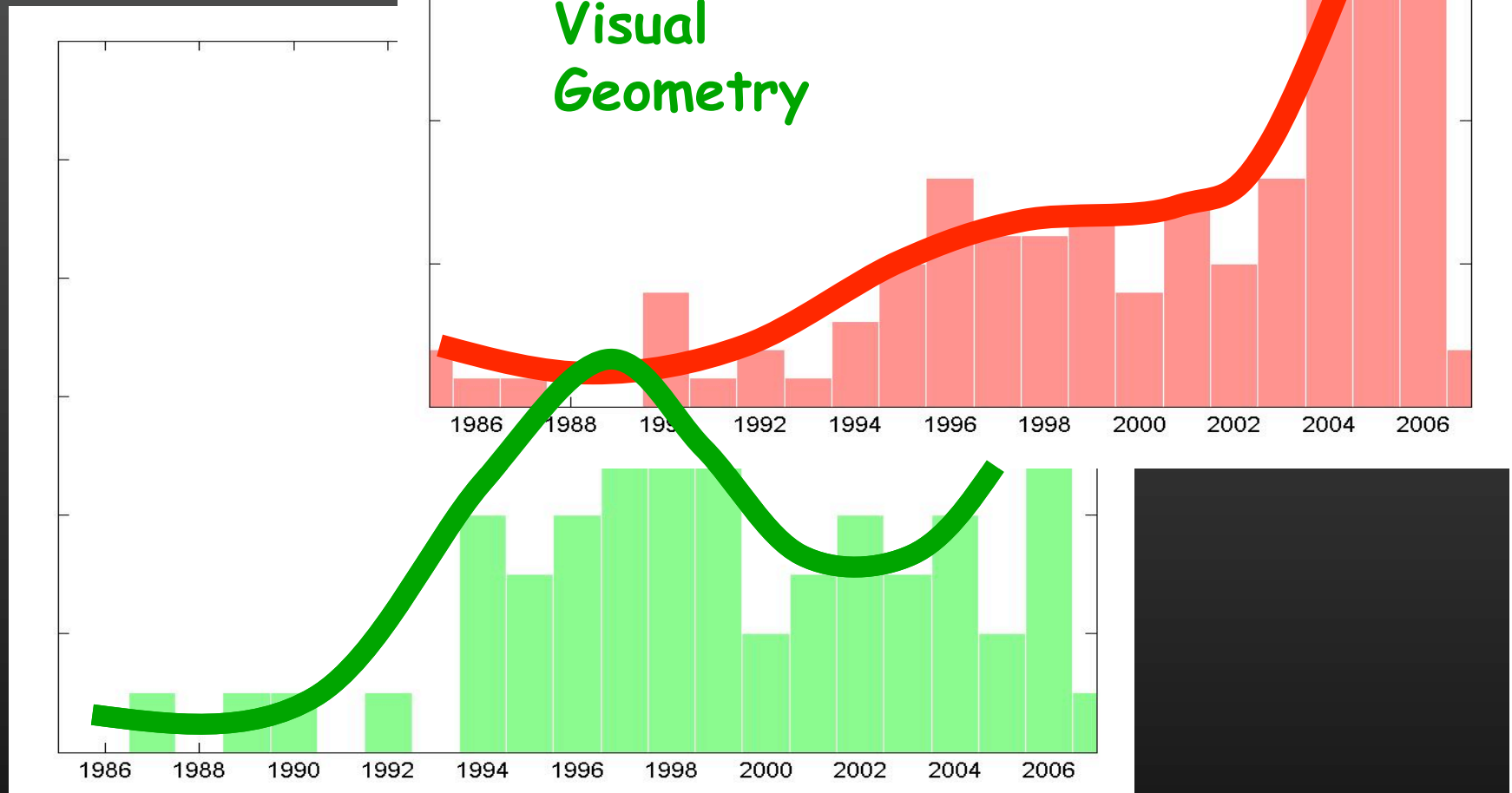


Felzenszwalb, McAllester, Ramanan (2007)
[Wins on 6 of the Pascal'07 classes, see Chum
& Zisserman (2007) for the other big winner.]

Number of research papers with
key-words "object recognition",
source: Springer.com



Numbers of papers
with key-words
"epipolar geometry"
source:
Springer.com



Visual Geometry:

Problems: Camera calibration, 3D reconstruction,

Structure from motion, stereo vision, ...

Scale



atching, ...

c.

Outline

- What computer vision is about
- What this class is about
- A brief history of visual recognition
- A brief recap on geometry -> J. Sivic

