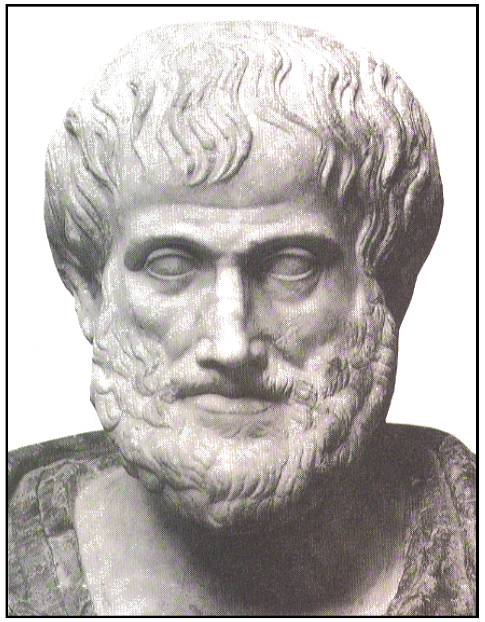
Introduction to Computational Vision

David J Fleet
Allan D Jepson

University of Toronto

for CSC 2503

What does it mean to see?

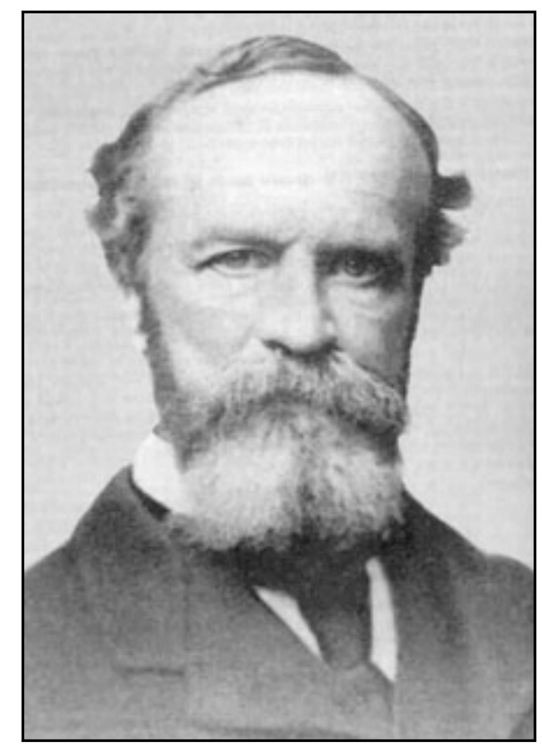


"To know what is where by looking."

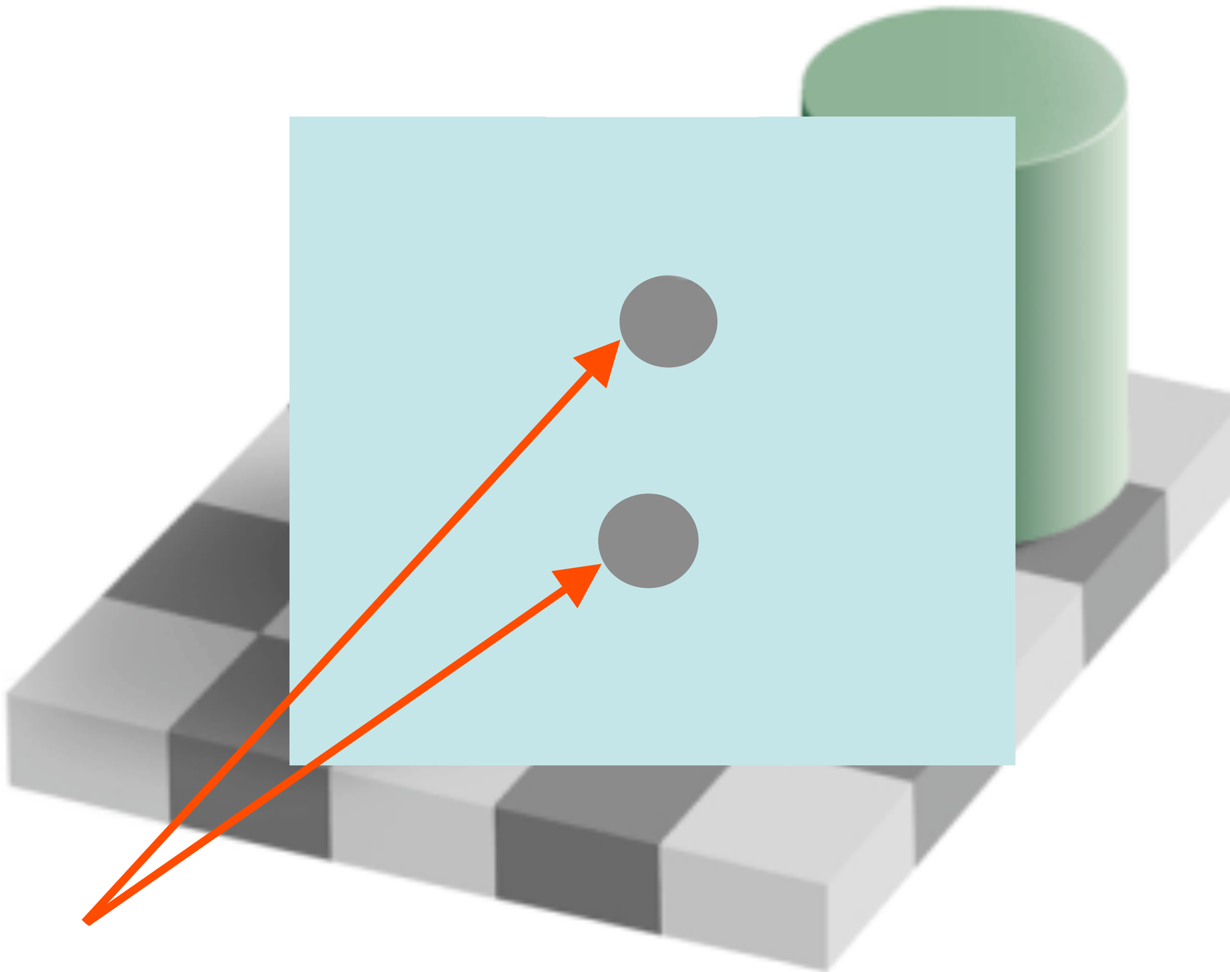
– Aristotle (300BC)

"Whilst part of what we perceive comes through our senses from the object before us, another part (and it may be the larger part) always comes out of our own mind."

– William James (1842-1910)



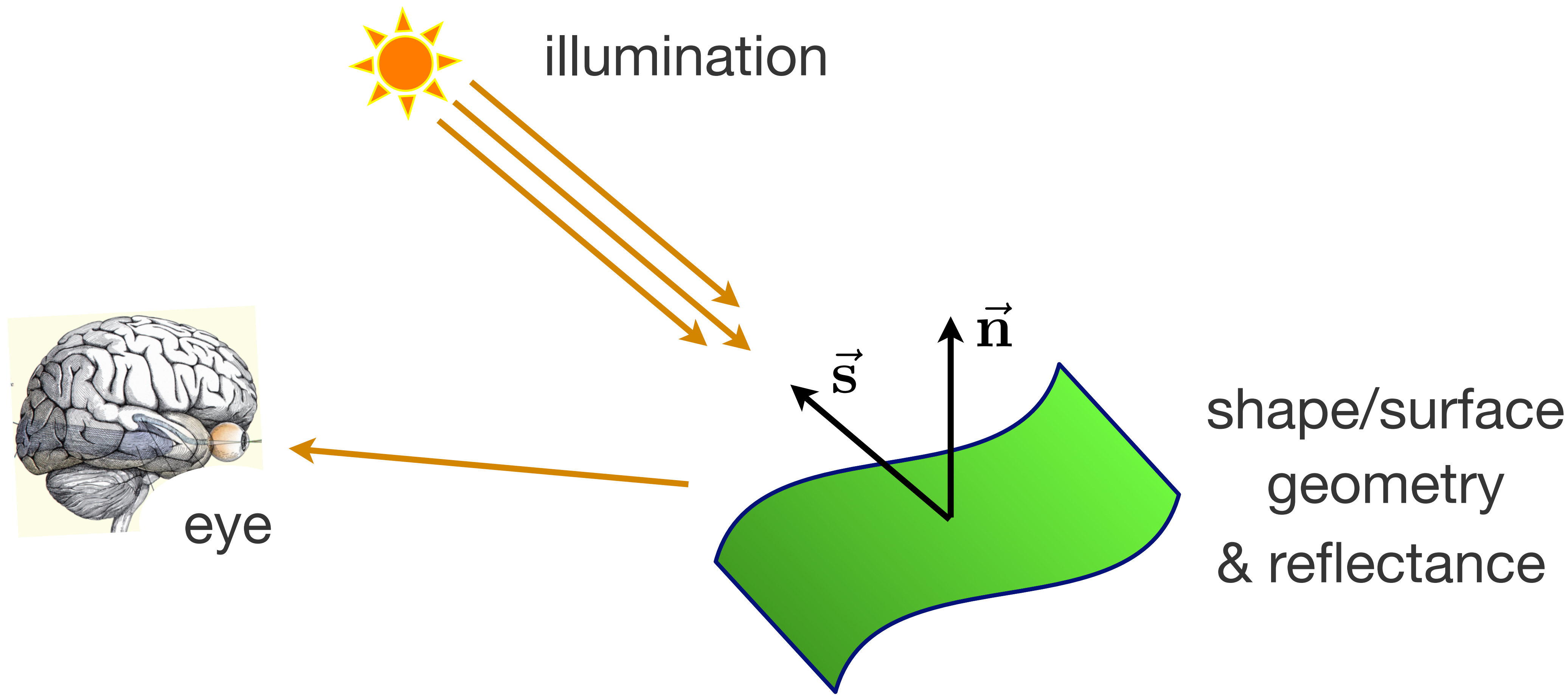
Lighting and appearance



same grey-level?
or different?

Edward H Adelson

Elements



Lambertian Model

$$L = R \times H(\vec{\mathbf{n}}, \vec{\mathbf{s}}, I)$$

reflected light

surface reflectance (albedo)

incident light

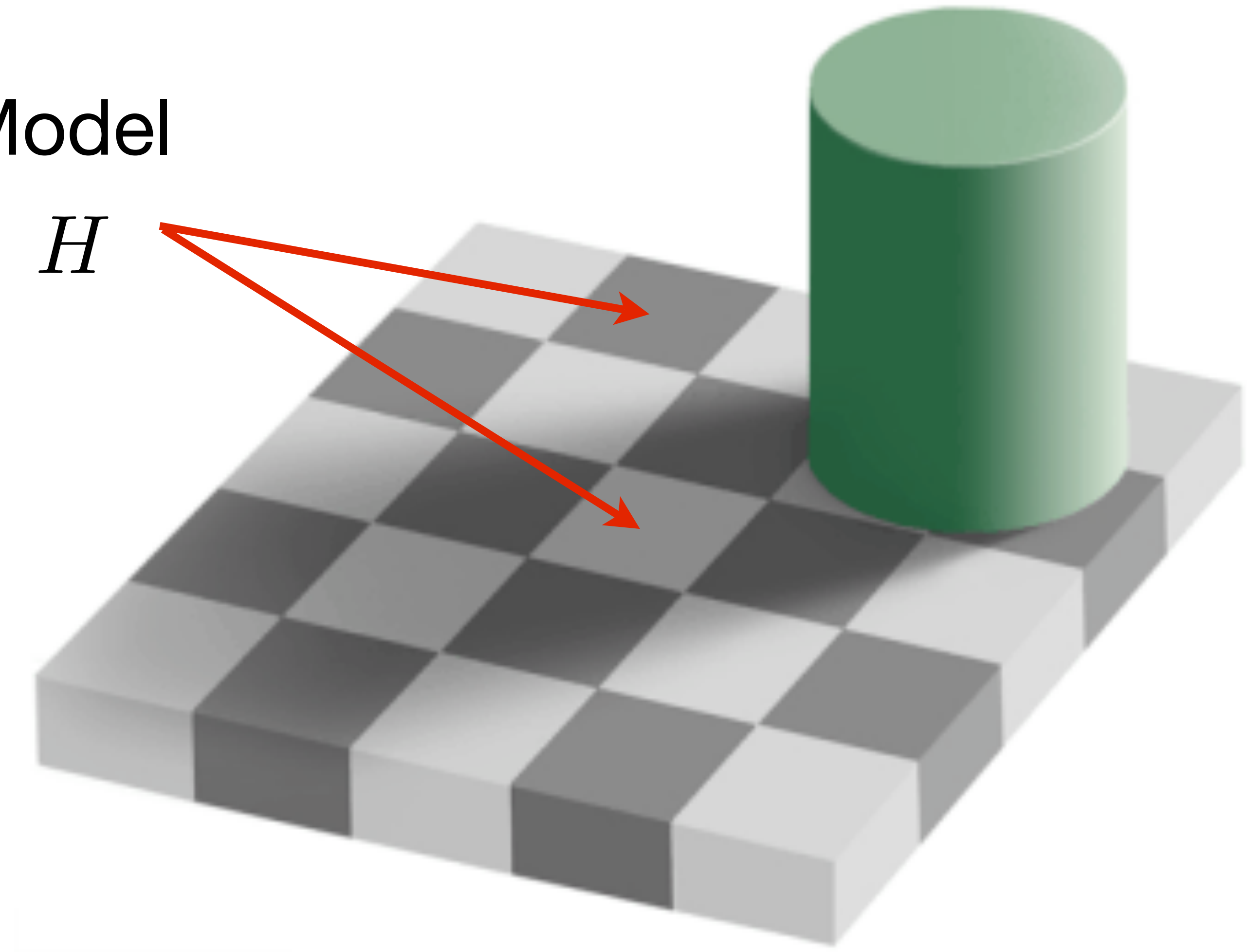
surface normal

light source direction & intensity

Lighting and appearance

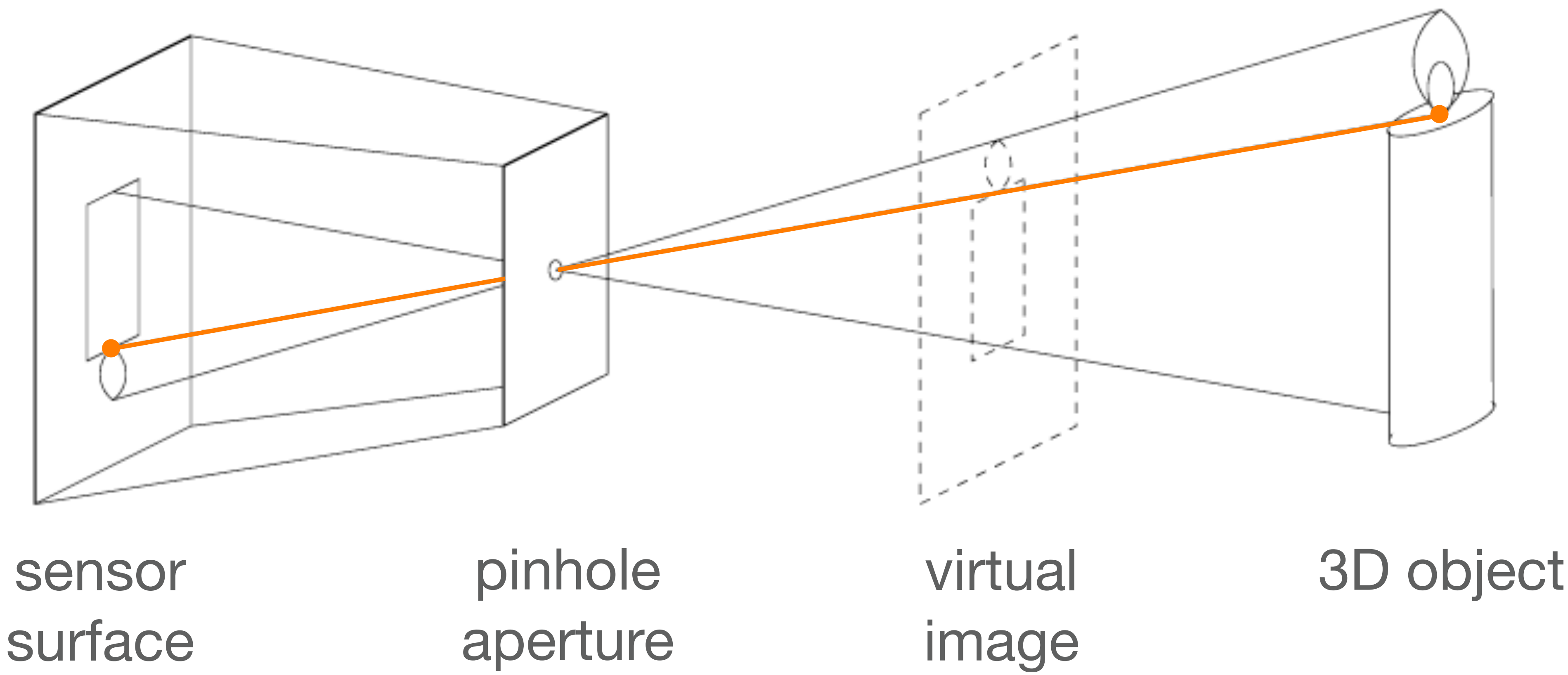
Lambertian Model

$$L = R \times H$$



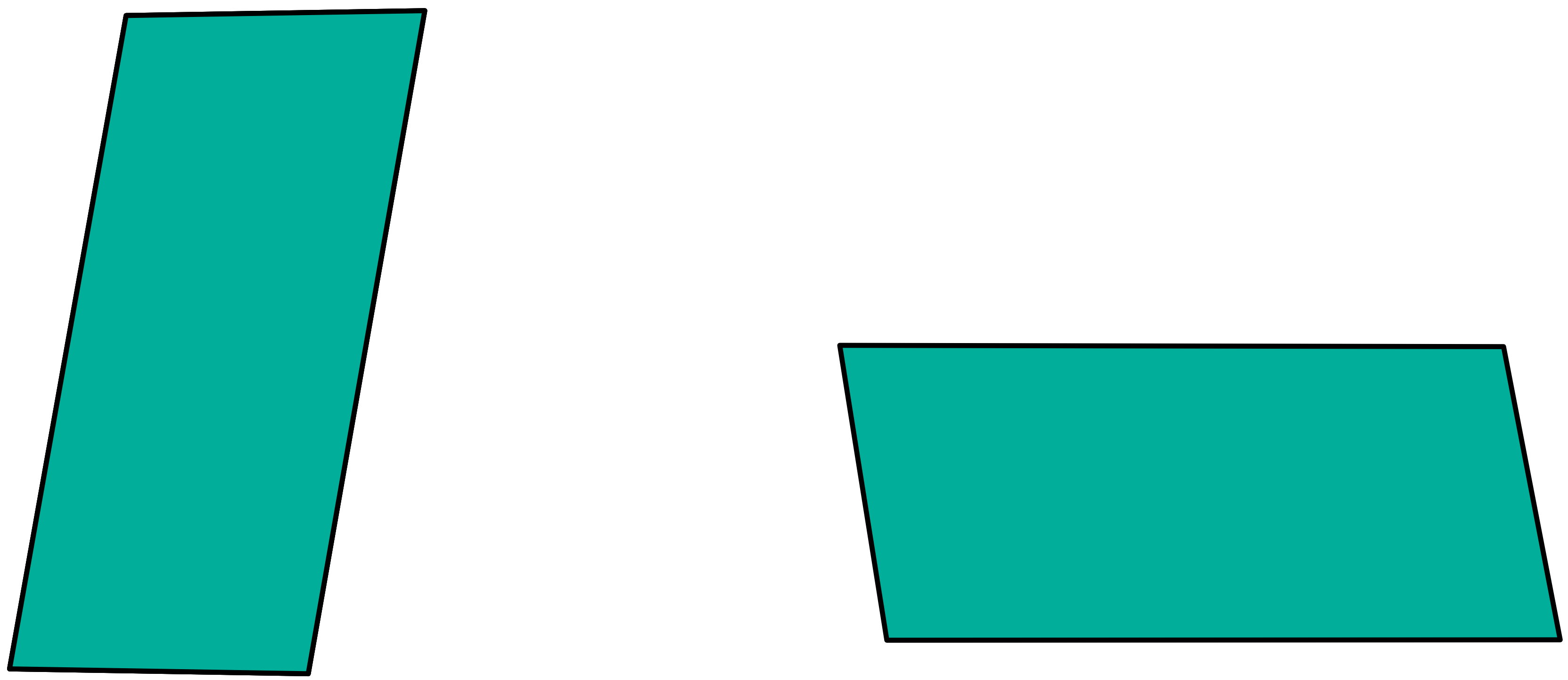
Edward H Adelson

Perspective projection



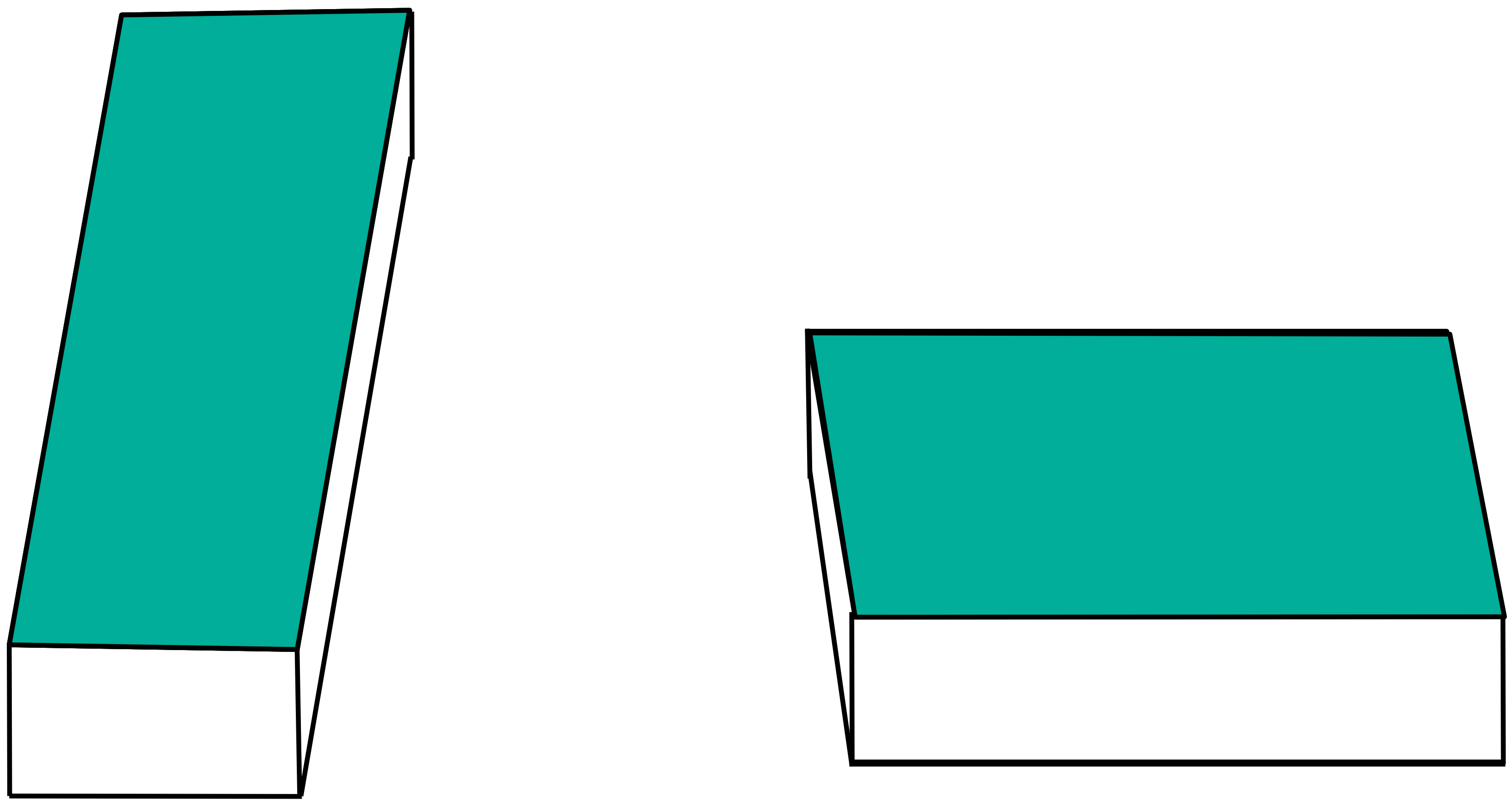
- The projection from 3D points onto the 2D image surface is modeled by perspective projection.
- Depth and size is lost in projection.

Objects are expected to be three-dimensional



Roger Shepard

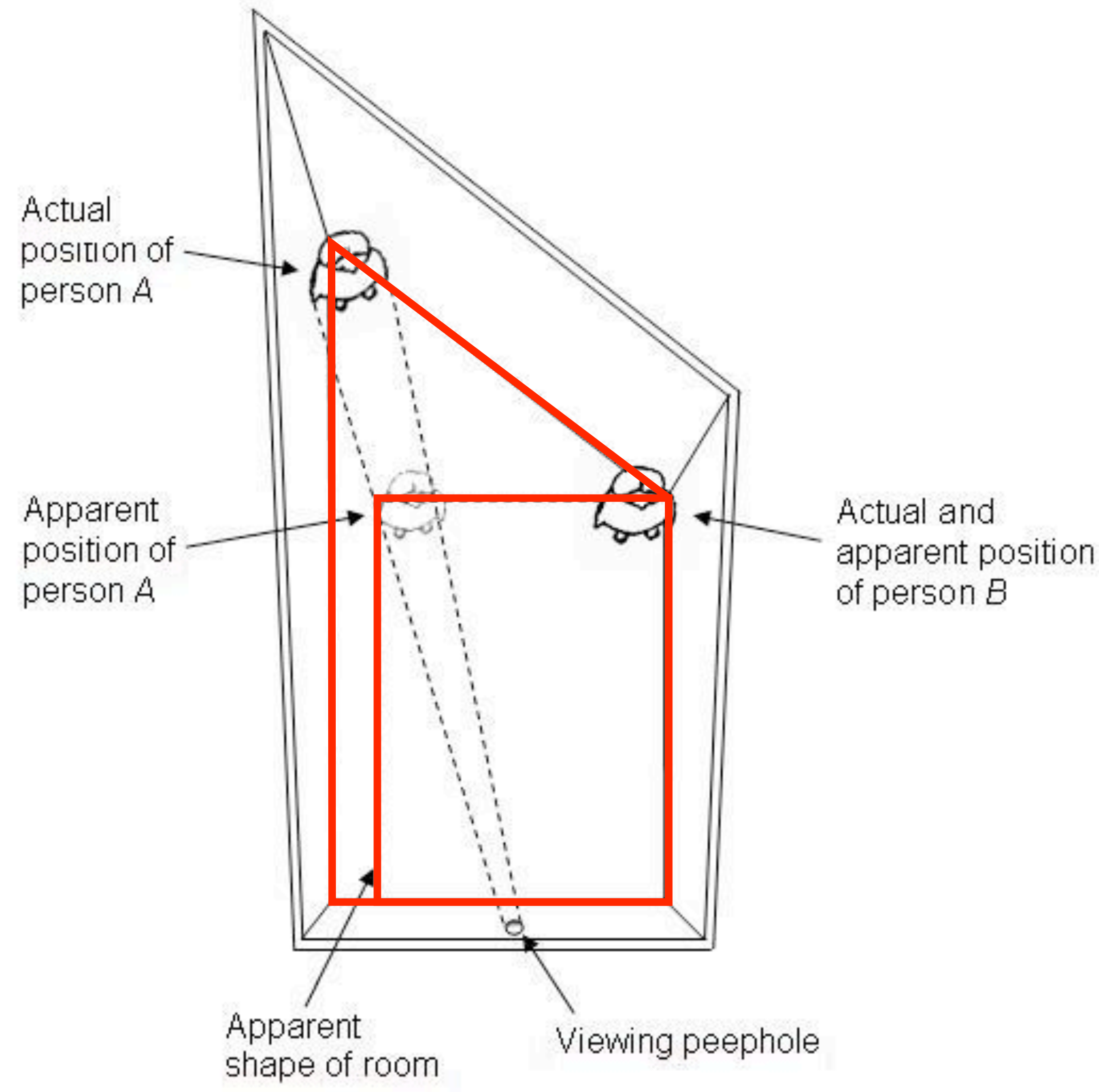
Objects are expected to be three-dimensional



Perception concerns properties of
3D objects, not the image per se.

Roger Shepard

Three-dimensional scene understanding



(Ames Room, Adelbert Ames, 1946)

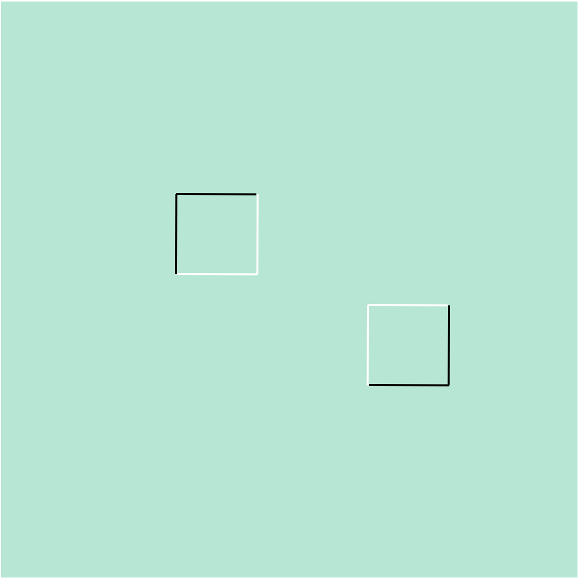
Local consistency and generic views

Jerry Andrus

Illumination is assumed to come from above

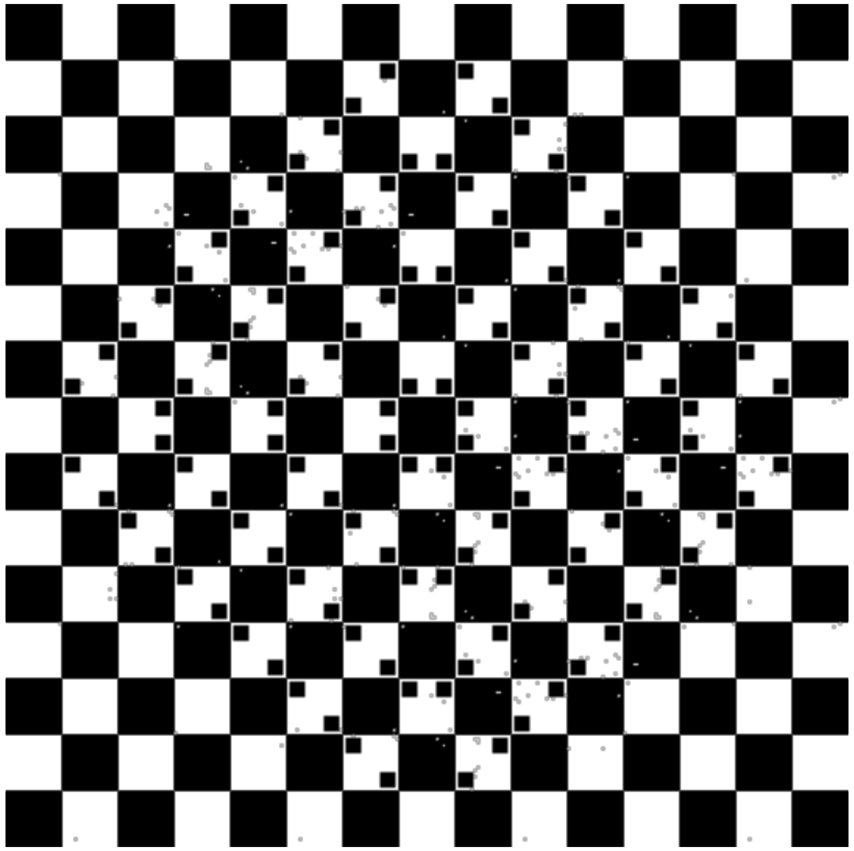
Walter Wick "Optical Tricks"

Illumination is assumed to come from above



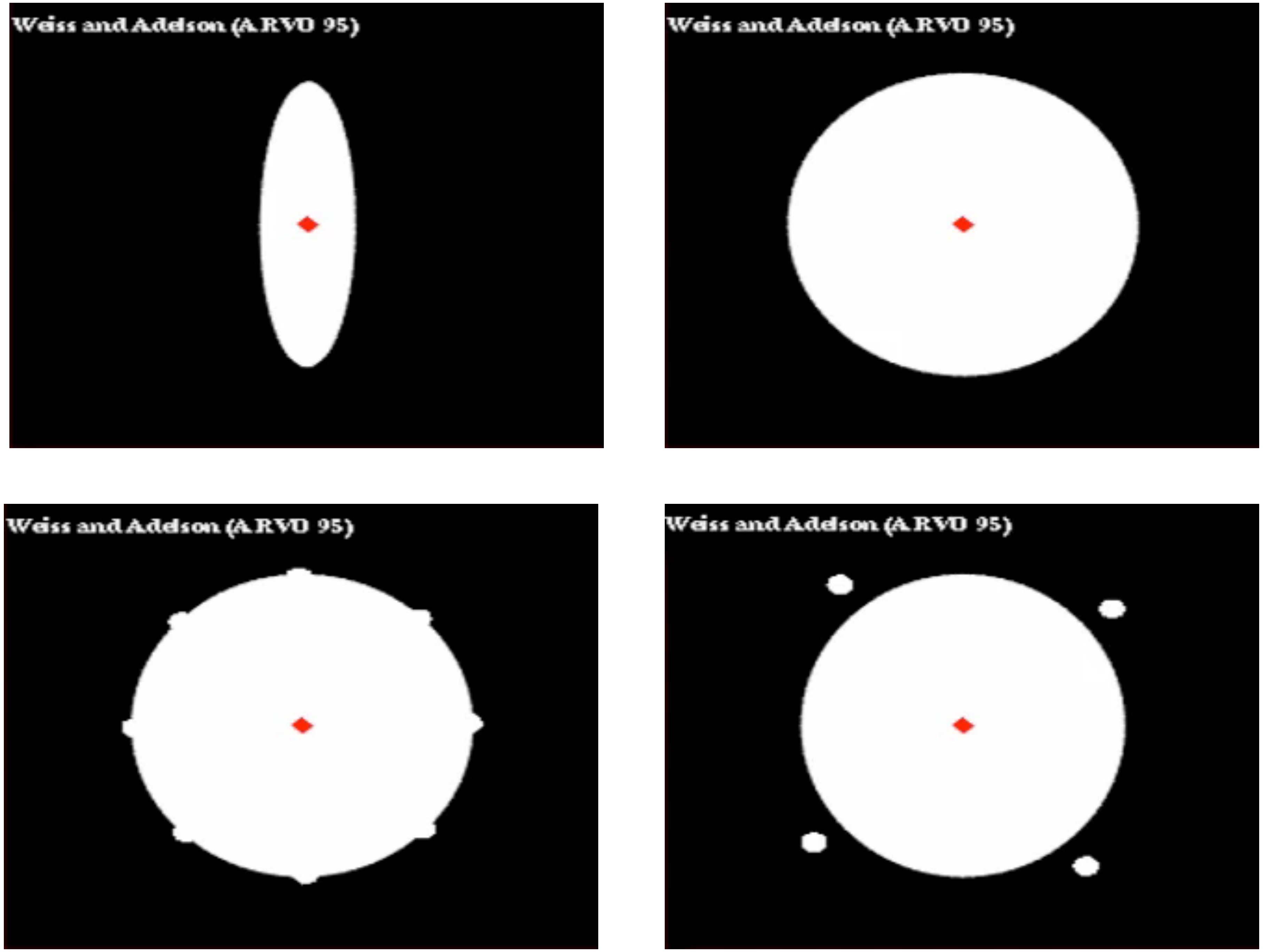
Niko Troje

Perceptual distortion



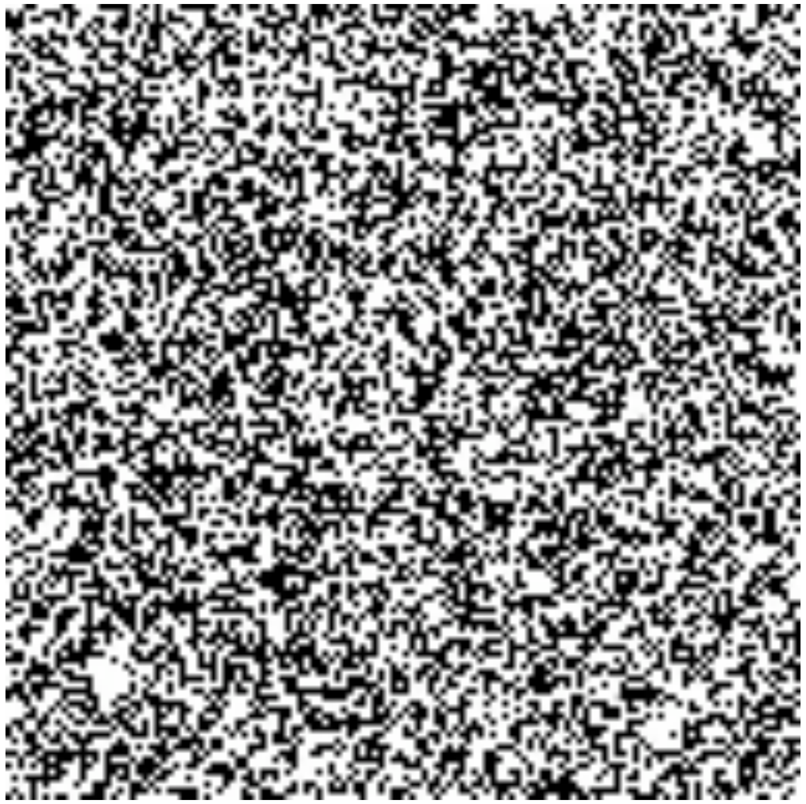
Akiyoshi Kitaoka

Smoothness and rigidity

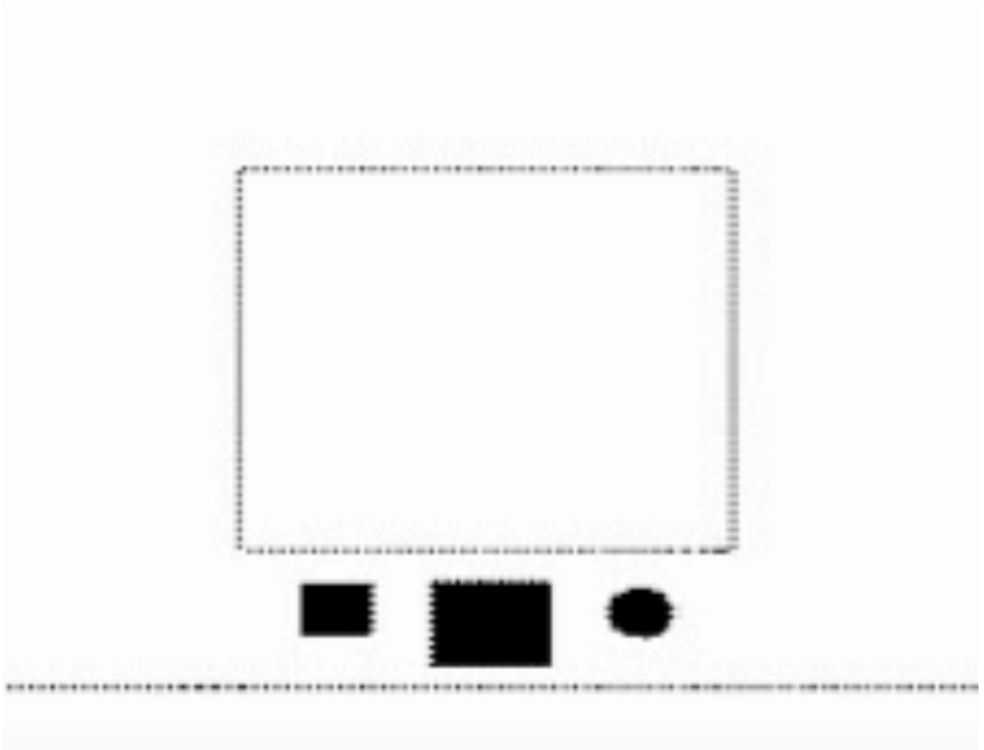


[Weiss and Adelson ’95]

Motion boundaries



Behaviour and intentions

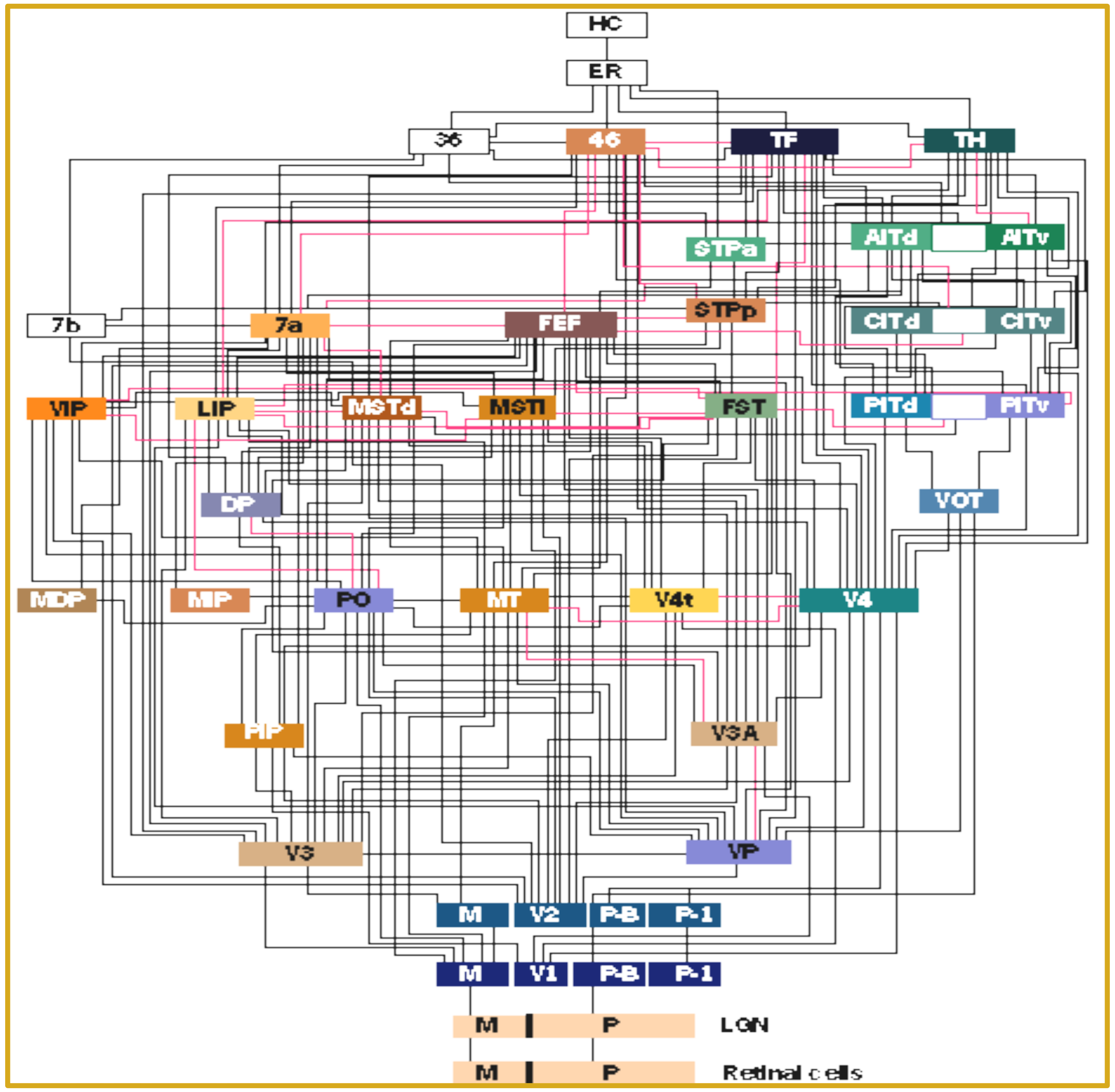


[Heider and Simmel '44]

Computational perception

Visual perception involves making inferences about the meaning of sensory data:

- surface properties
- illumination
- object size and depth



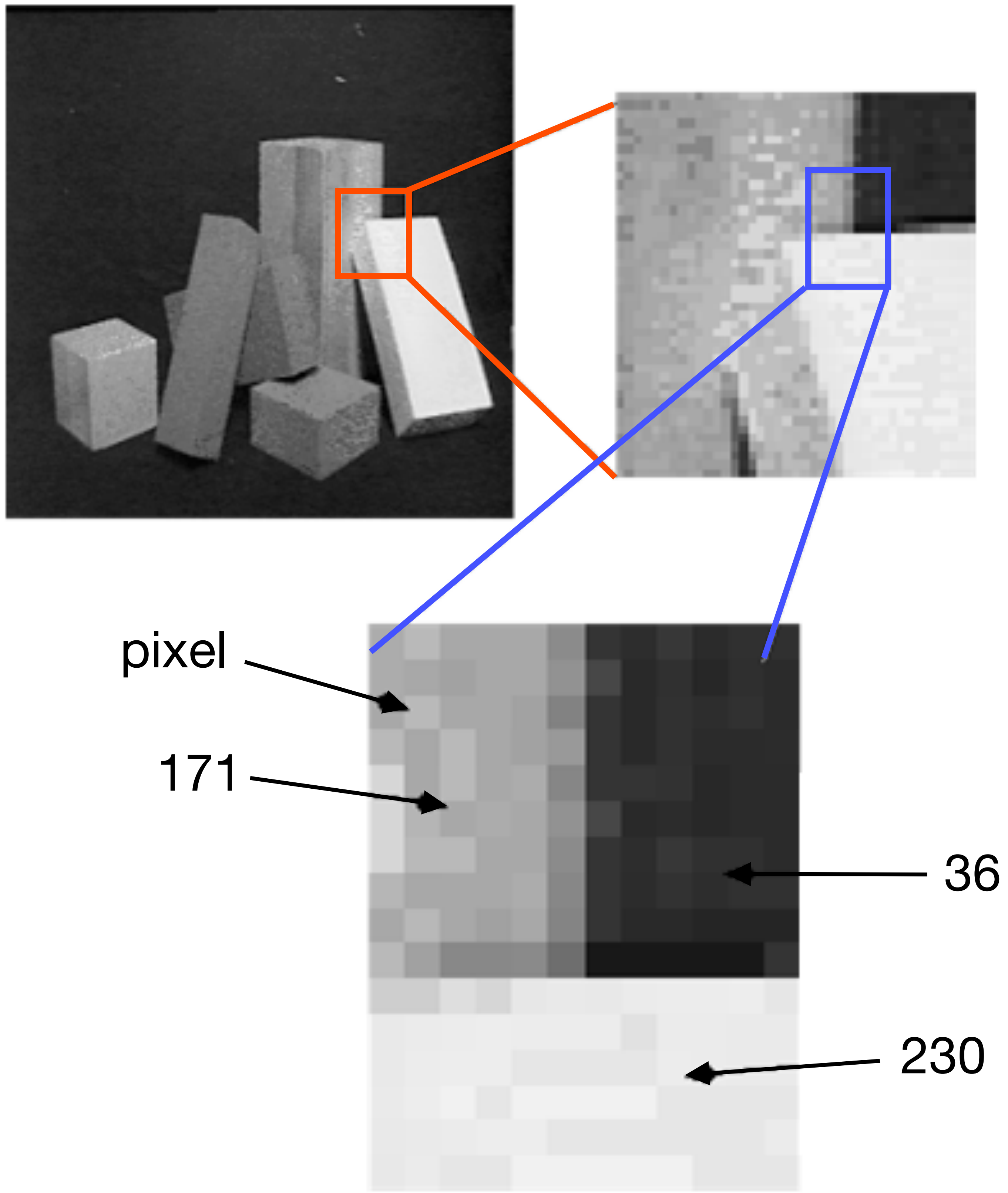
Computational perception entails the mathematical specification of such inference problems, along with algorithms to solve them.

Computational perception

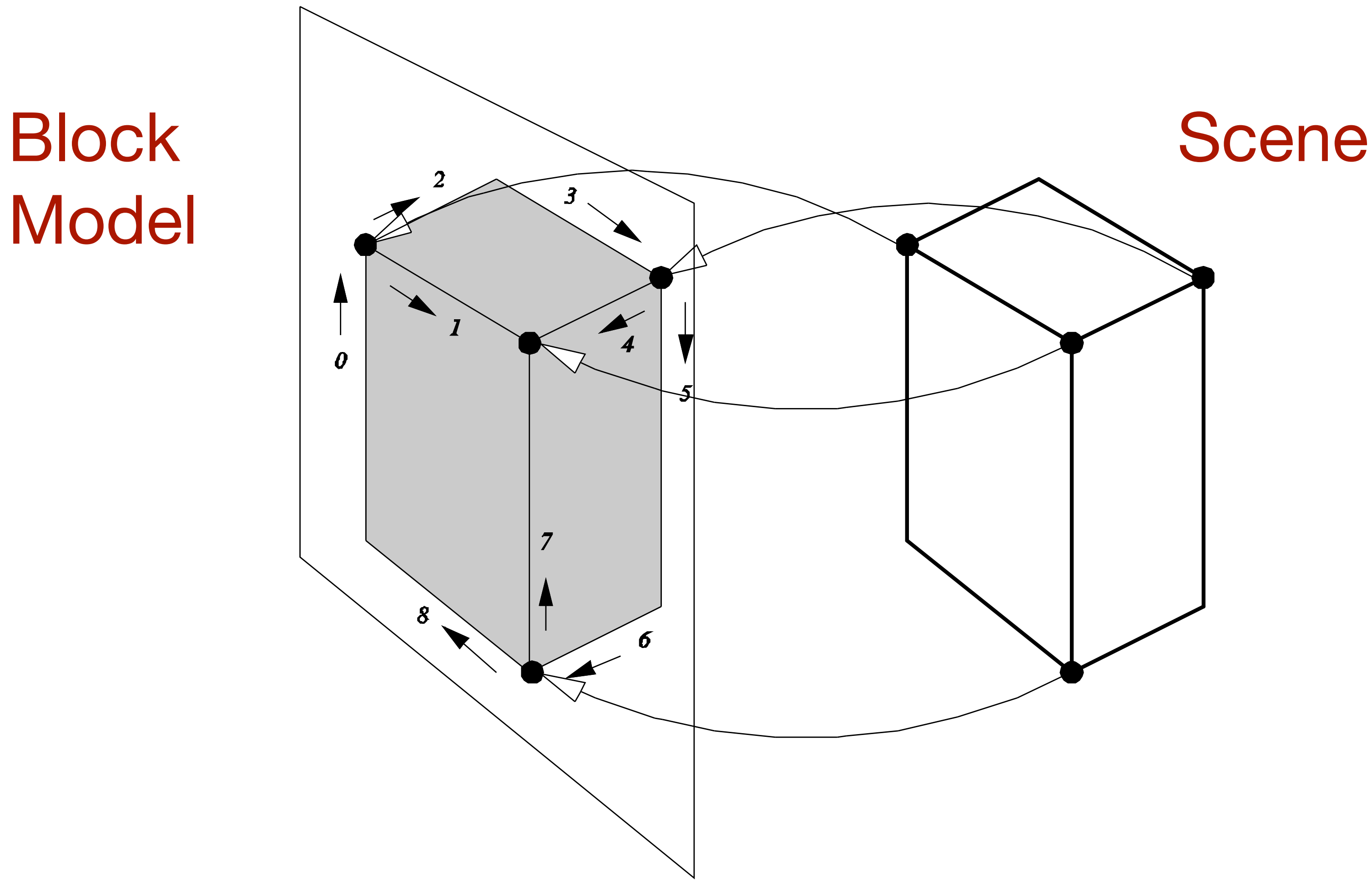
Key elements of computational perception (models):

- Scene domain theory
  (to specify model classes / parameters of interest)
- Measurement model
  (mapping from scenes to image measurements)
- Plausibility theory
  (measure the plausibility of “consistent” interpretations)
- Search
  (effective methods for finding best interpretations)

Blocks world example



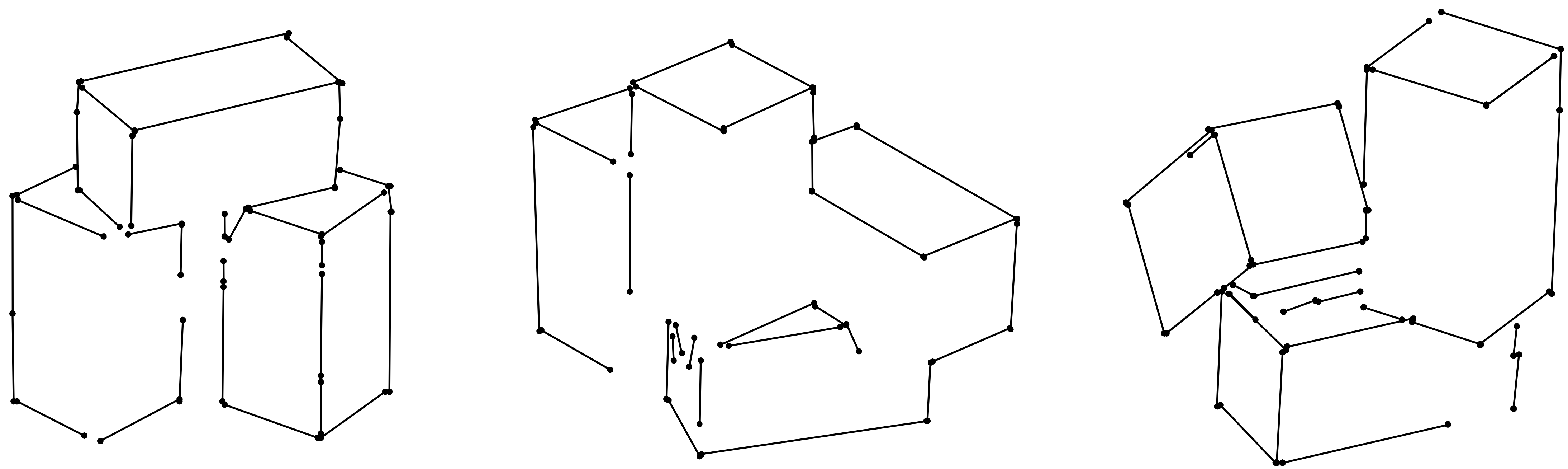
Blocks world domain theory



The model comprises blocks (arranged in depth layers) and sticks (isolated line segments).

Blocks are opaque, so they can occlude other objects.

Edge measurements

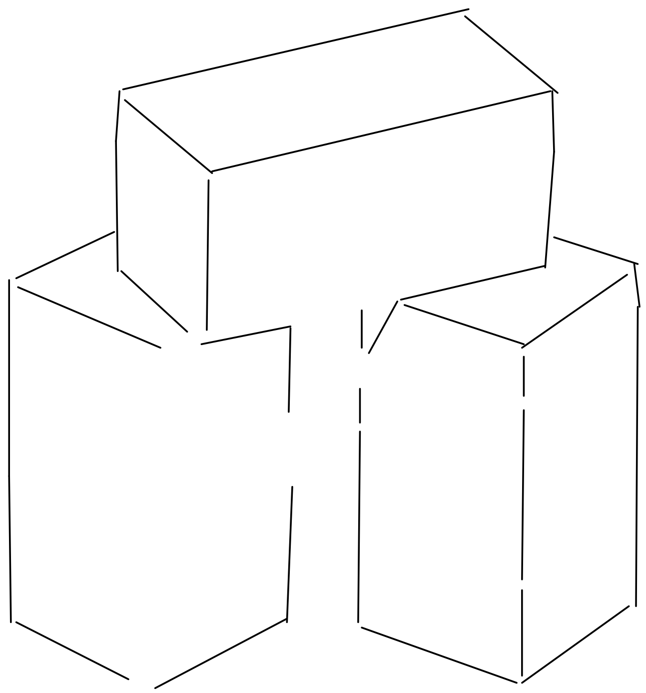


Edges convey useful information about blocks world scenes, but measurements are noisy due to photon noise and modeling error.

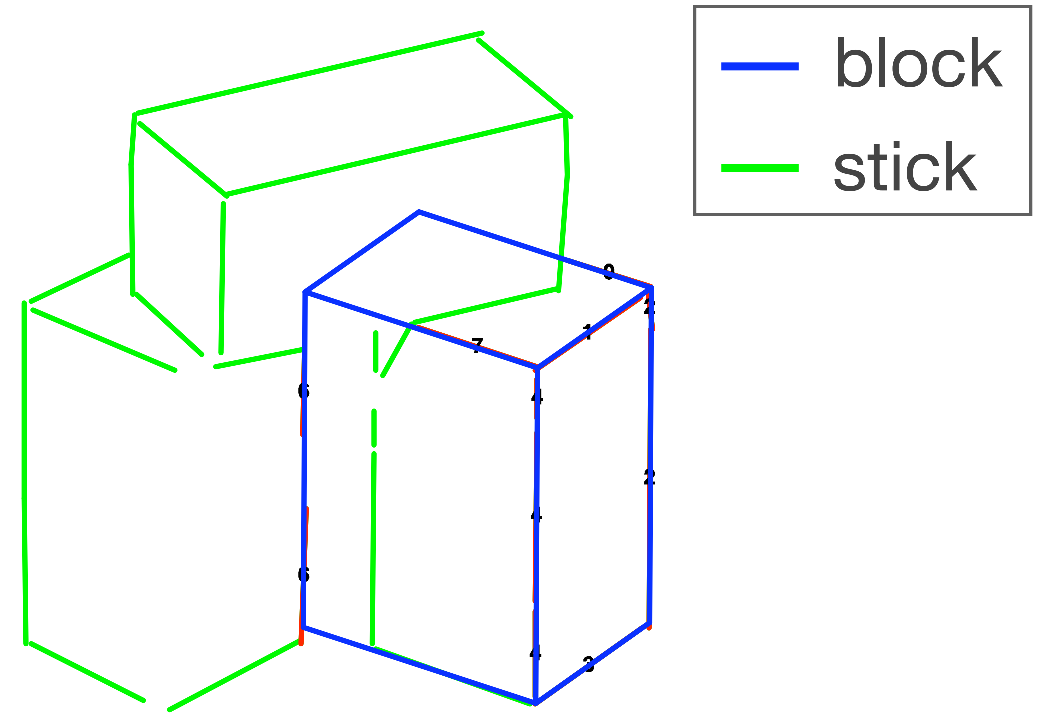
Edges are sometimes missing, or broken with missing fragments.

Image interpretations explain all edges in terms of blocks & sticks.

Consistent interpretations

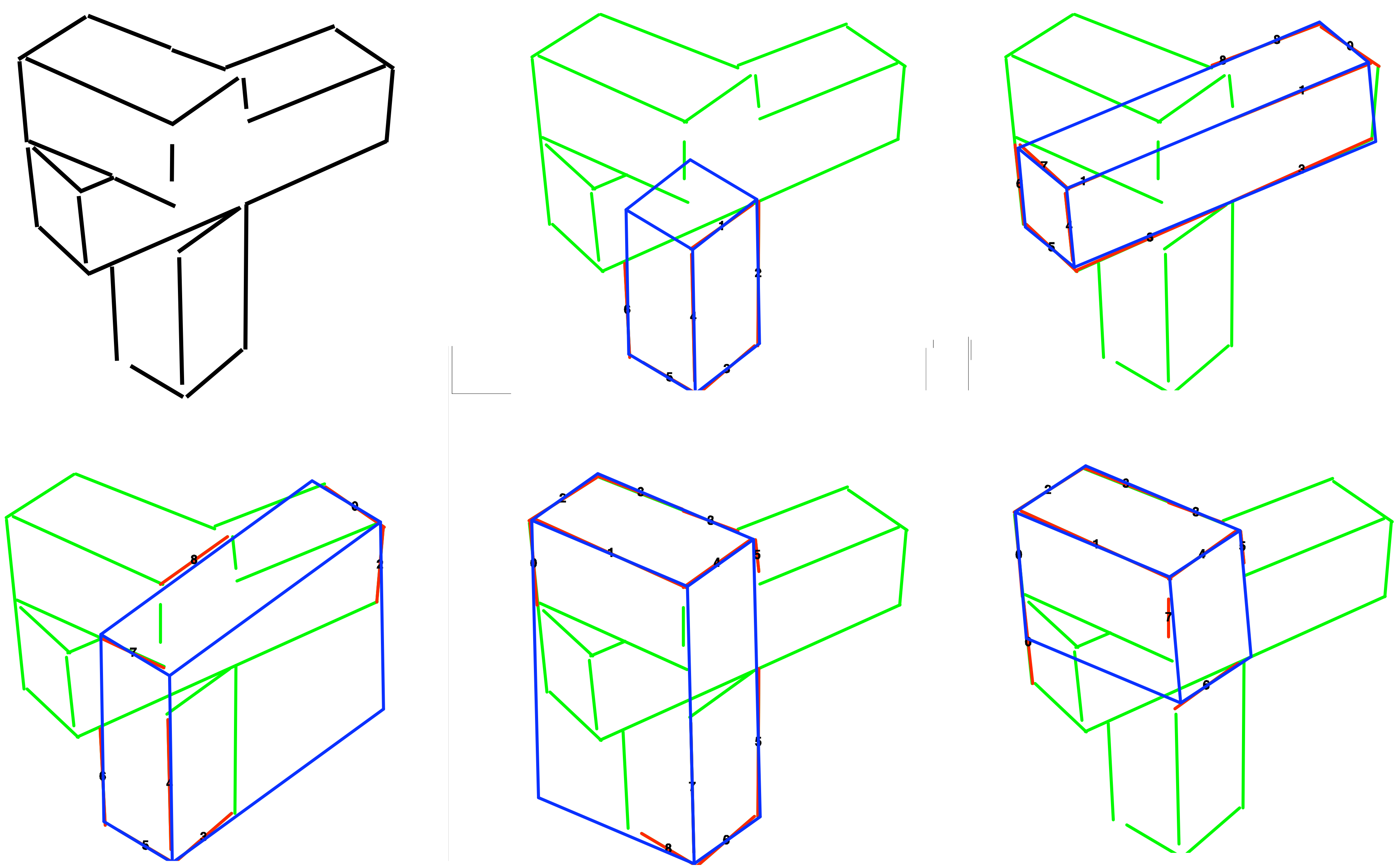


Edge segments
(34 edges here)



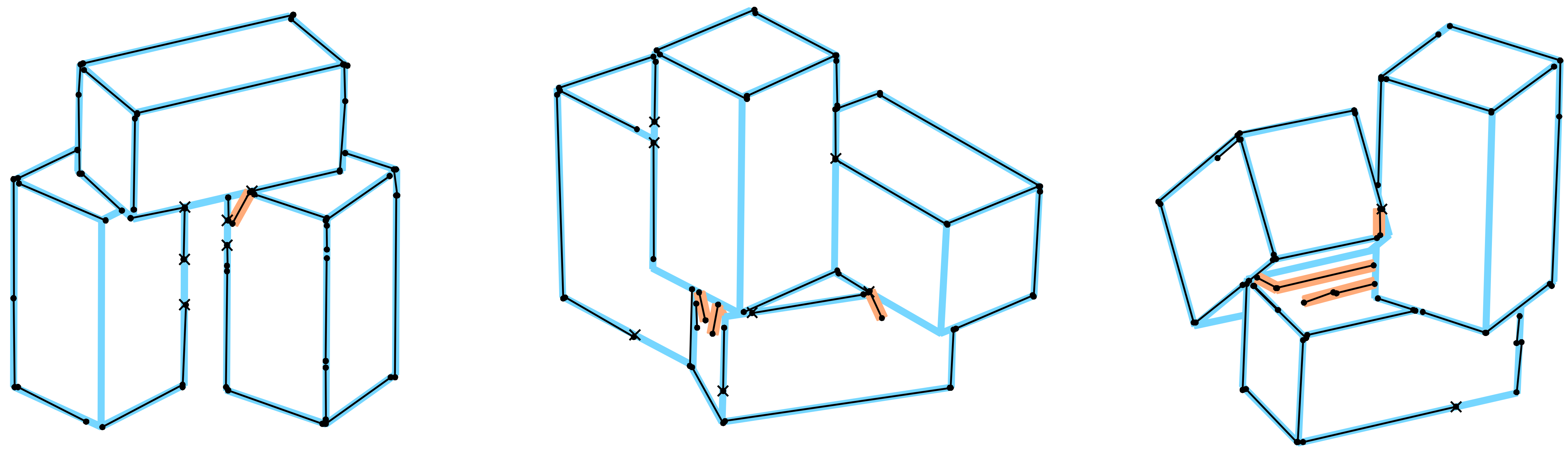
Candidate block
(interpretation)

Candidate blocks



The number of “consistent” blocks is large (4-6K).

Search for plausible interpretations



Search over plausible models, with a suitable plausibility measure.

Often, one model is overwhelmingly more plausible that all others.

How can we measure the plausibility of an interpretation,
and how do we search for the best interpretations?

Bayesian inference

Inference: Reasoning with uncertain beliefs according to a probabilistic calculus.

Beliefs characterized in terms of probability distributions over events (domain variables, models, …)

Thomas Bayes
(1702-1761)

"Probability theory is nothing more than common sense reduced to calculation."
– Pierre-Simon Laplace (1749-1827)

Bayes' Rule

Model Parameters: M

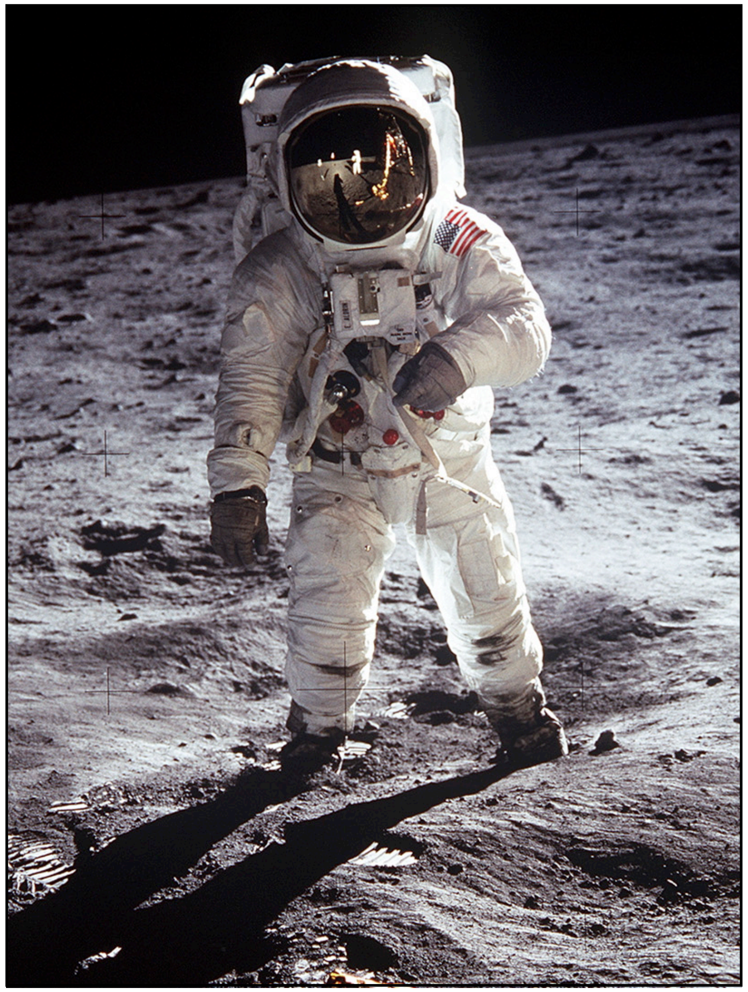
Data (Observations): D

$$\underbrace{p(M \mid D)}_{\text{posterior}} = \frac{\overbrace{p(D \mid M)}^{\text{likelihood}} \; \overbrace{p(M)}^{\text{prior}}}{p(D)}$$

Looking at People

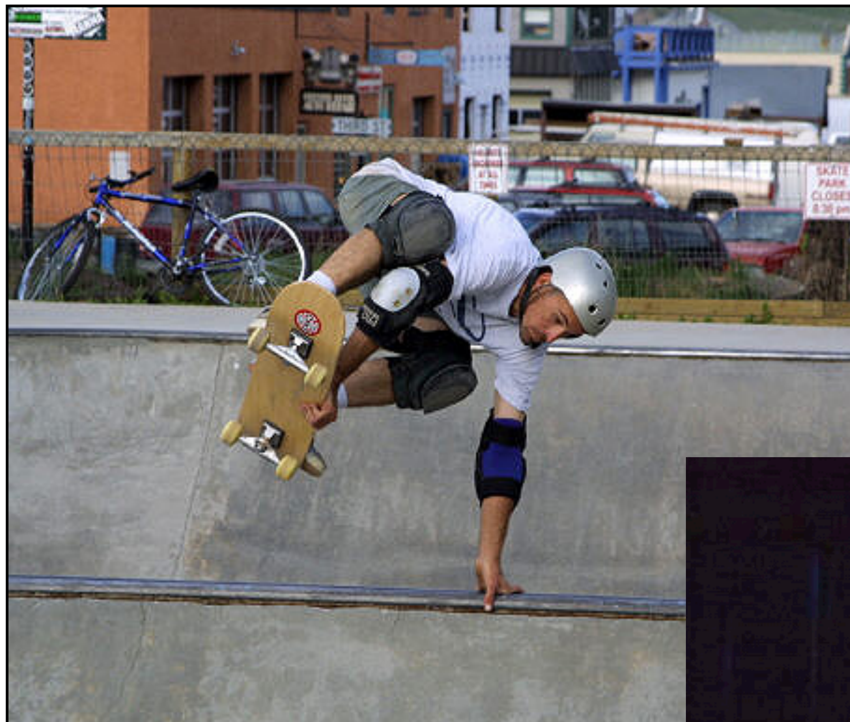
- detect and recognize people
- estimate pose and motion
- recognize gestures and actions

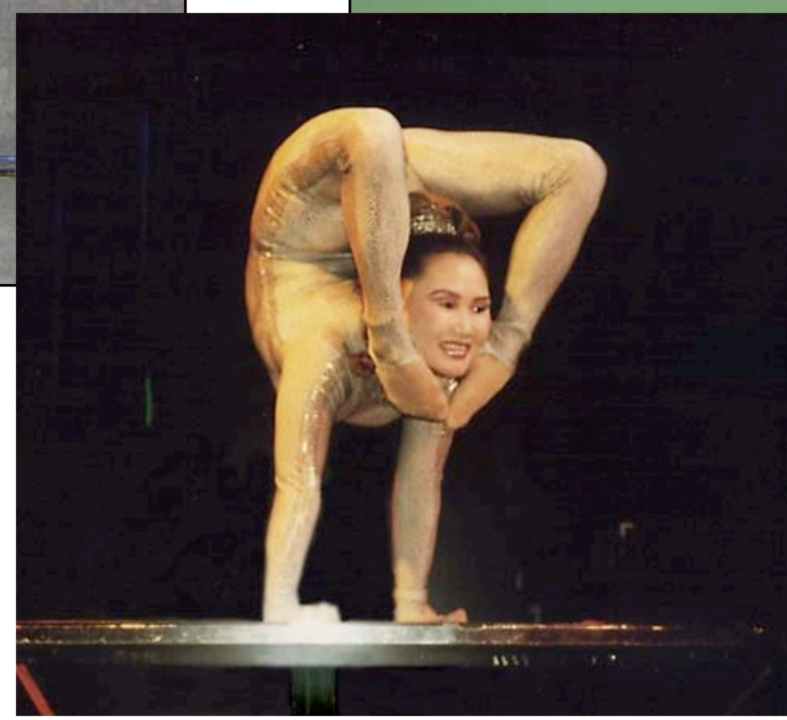
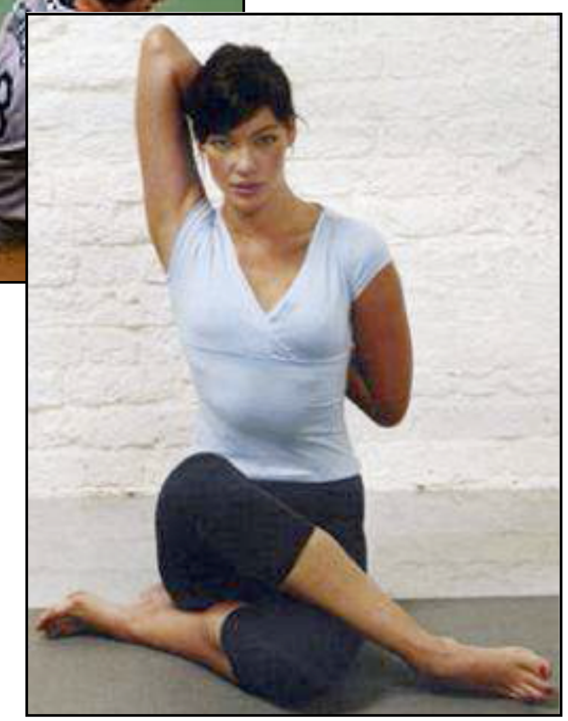
Challenges: Appearance, size and shape



People come in all shapes and sizes, with highly variable appearance.

Challenges: Complex pose / motion



People have many degrees of freedom, comprising an articulated skeleton overlaid with soft tissue and deformable clothing.

Challenges: Complex movements

People move in complex ways, often communicating with subtle gestures

Challenges: Noisy and missing measurements

Ambiguities in pose are commonplace, due to

- background clutter
- apparent similarity of parts
- occlusions
- loose clothing
- ...

Challenges: Appearance variability

Image appearance changes dramatically over time
due to non-rigidity of body and clothing and lighting.

Challenges: Appearance variability

Image appearance changes dramatically over time due to non-rigidity of body and clothing and lighting.

Challenges: Context dependence

Perceived scene context influences object recognition.

[Courtesy of Antonio Torralba]

Conclusion

For most computer vision problems we face similar issues:

- What are the models and parameters that we want to estimate?
- What are the informative image measurements?
- How do we select specific models given the measurements?
- How to we search this space of models/parameters efficiently?

Current practice – simple models:

- Small sets of known objects with specific appearance and/or form (e.g., human faces, cars, …)
- “lower-level” measurement of image and scene properties (e.g., motion depth, …)

This course aims to introduce you to the fundamentals and the current practice, and to prepare you for research in computer vision.