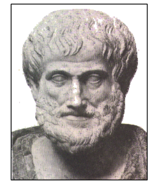
Introduction to Computational Vision

David J Fleet and Allan$_+$ Jepson

University of Toronto

What does it mean to see?

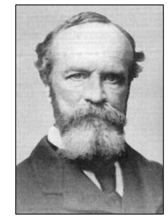


"To know what is where by looking."

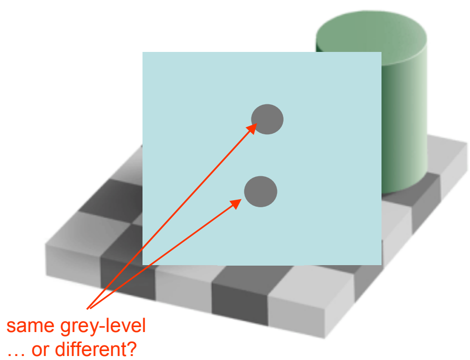
– Aristotle (300BC)

"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)

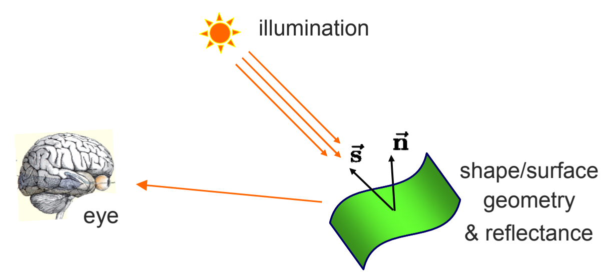


Lighting and appearance



Edward H Adelson

Elements



Lambertian Model

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

reflected light

surface reflectance (albedo)

incident light on surface

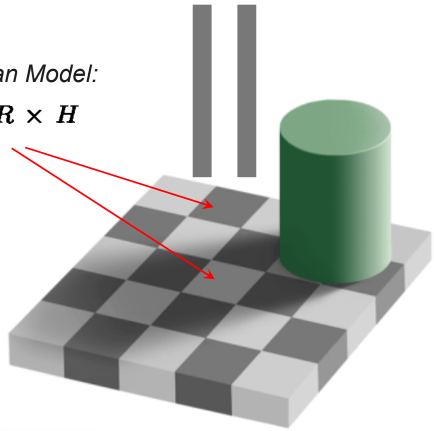
surface normal

light source direction & intensity

Lighting and appearance

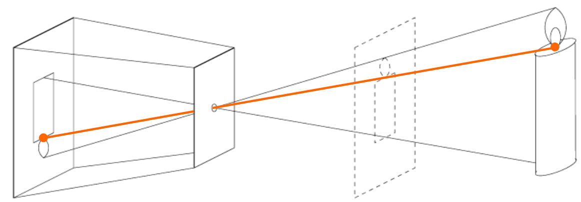
Lambertian Model:

$$L = R \times H$$



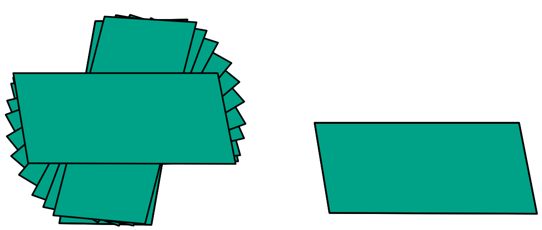
Edward H Adelson

Geometry: Pinhole camera model



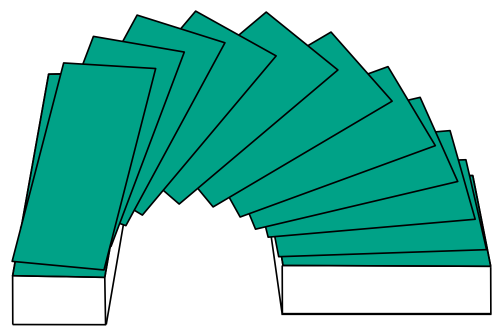
- The projection from 3D points onto the 2D image surface is modeled by *perspective projection.*
- Depth and size is lost in projection.

Three-dimensional scene inference



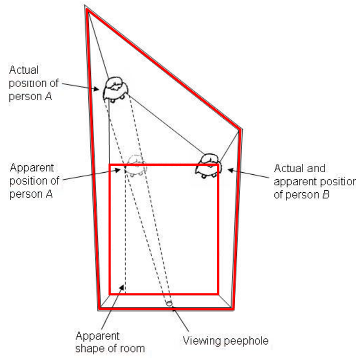
Roger Shepard

Three-dimensional scene inference



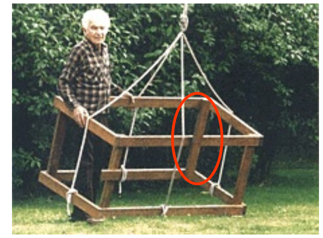
Roger Shepard

Three-dimensional scene inference



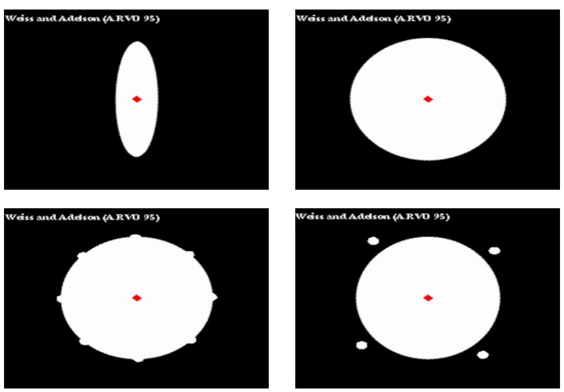
(*Ames Room*, Adelbert Ames, 1946)

Local consistency and generic views



Jerry Andrus

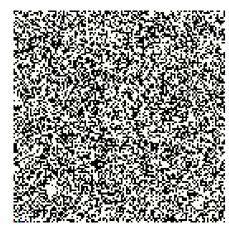
Motion: smoothness or rigidity



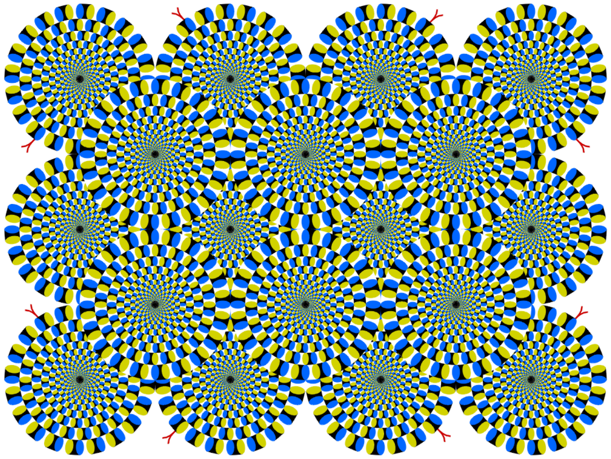
Sometimes we prefer slow and smooth motions over rigid interpretations?

Weiss and Adelson (1995)

Motion-defined form

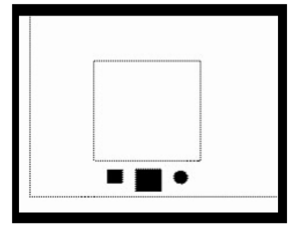


Form Suggests Motion: Rotating Snakes Illusion



Akiyoshi Kitaoka

Inference of behaviour and intentions

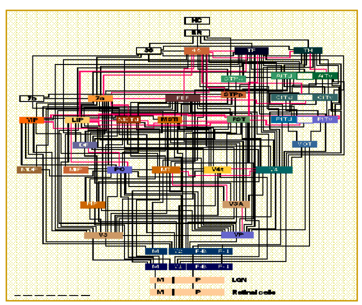


Heider and Simmel (1944)

Computational perception

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

- surface properties
- illumination
- motion
- object size and depth



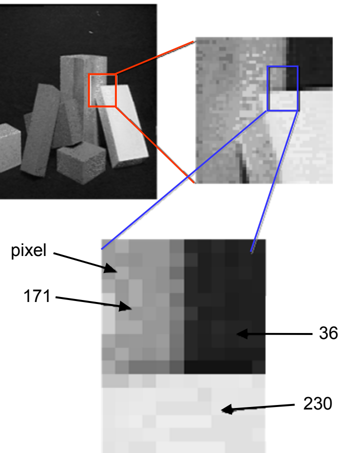
Computational perception involves the mathematical specification of such inference problems, along with algorithms to solve them.

Theories

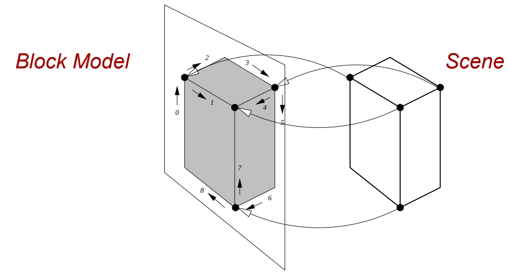
Key elements of computational perception (*created by us*):

- 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



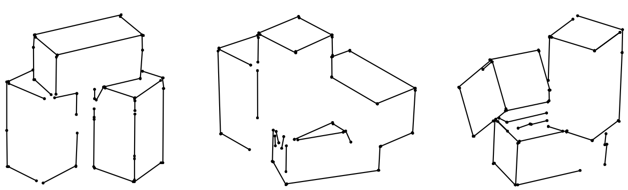
Blocks world domain theory



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

Blocks are opaque, so they can occlude other objects.

Edge measurements

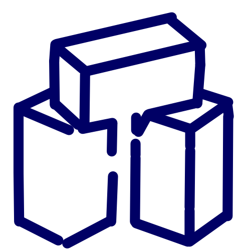


Edges convey useful information about blocks world scenes, but measurements are noisy due to photon noise and modeling error.

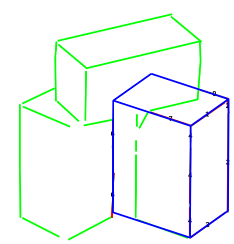
So, edges are sometimes missing, or broken with missing fragments.

Image interpretations must explain all edges in terms of blocks or sticks.

Consistent interpretation

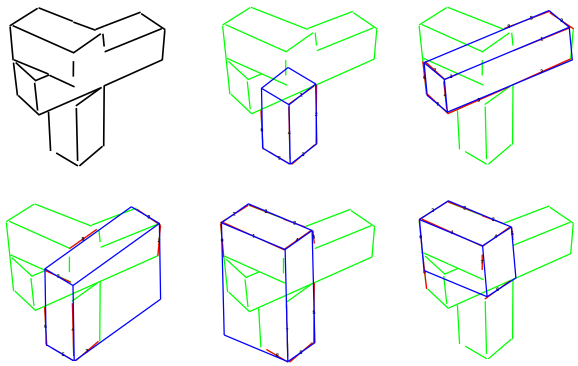


Edge segments
(34 edges here)



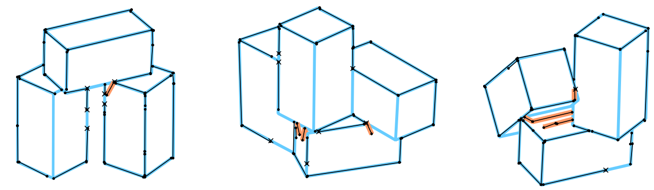
Candidate block
(interpretation)

Candidate blocks



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

Search for optimal interpretations



Search over plausible models, with a suitable measure of plausibility.
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)

Inference

Model Parameters: M
Data (Observations): D

Likelihood
Prior
Posterior

$$p(M \mid D) = \frac{p(D \mid M)\ p(M)}{p(D)}$$

Looking at people

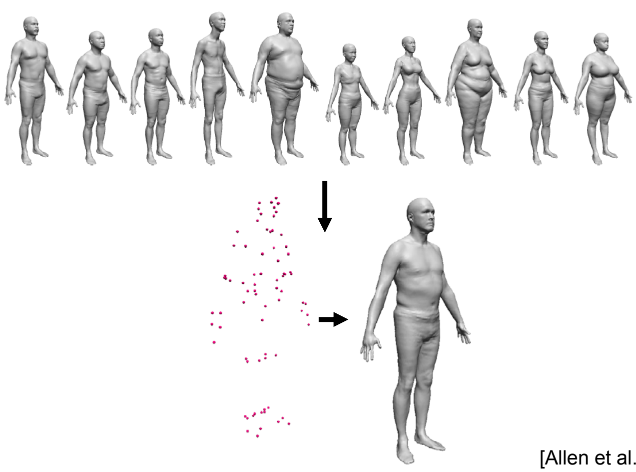
Tasks:

- Detect and recognize people
- Estimate pose, motion, and shape
- Recognize gestures and actions

Challenges: Appearance, size and shape

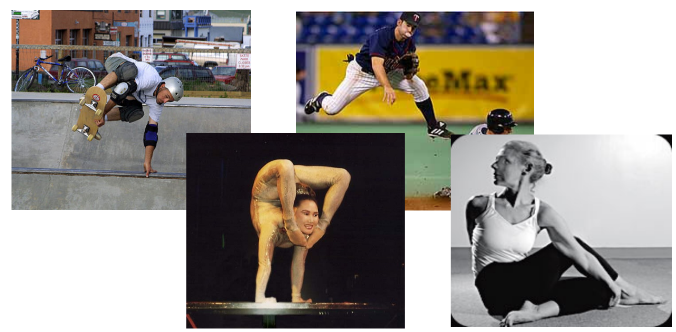
People come in all shapes and sizes, with highly variable appearance.

Shape representation and inference



[Allen et al. 2003]

Challenges: Complex pose / motions



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

Challenges: Complex movements

Silly walks

Social display of puzzlement

People move in complex ways and often convey information with subtle gestures.

Challenges: Resolution, occlusion, backgrounds

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

Appearance changes dramatically over time due to non-rigidity of body, lighting and occlusions.

Conclusions

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 do 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 further graduate work in computer vision.