CSC2503: Foundations of Computational Vision

Kyros Kutulakos

University of Toronto

http://www.cs.toronto.edu/~kyros/courses/2503

The Summer Vision Project

Download

Author: Papert, Seymour

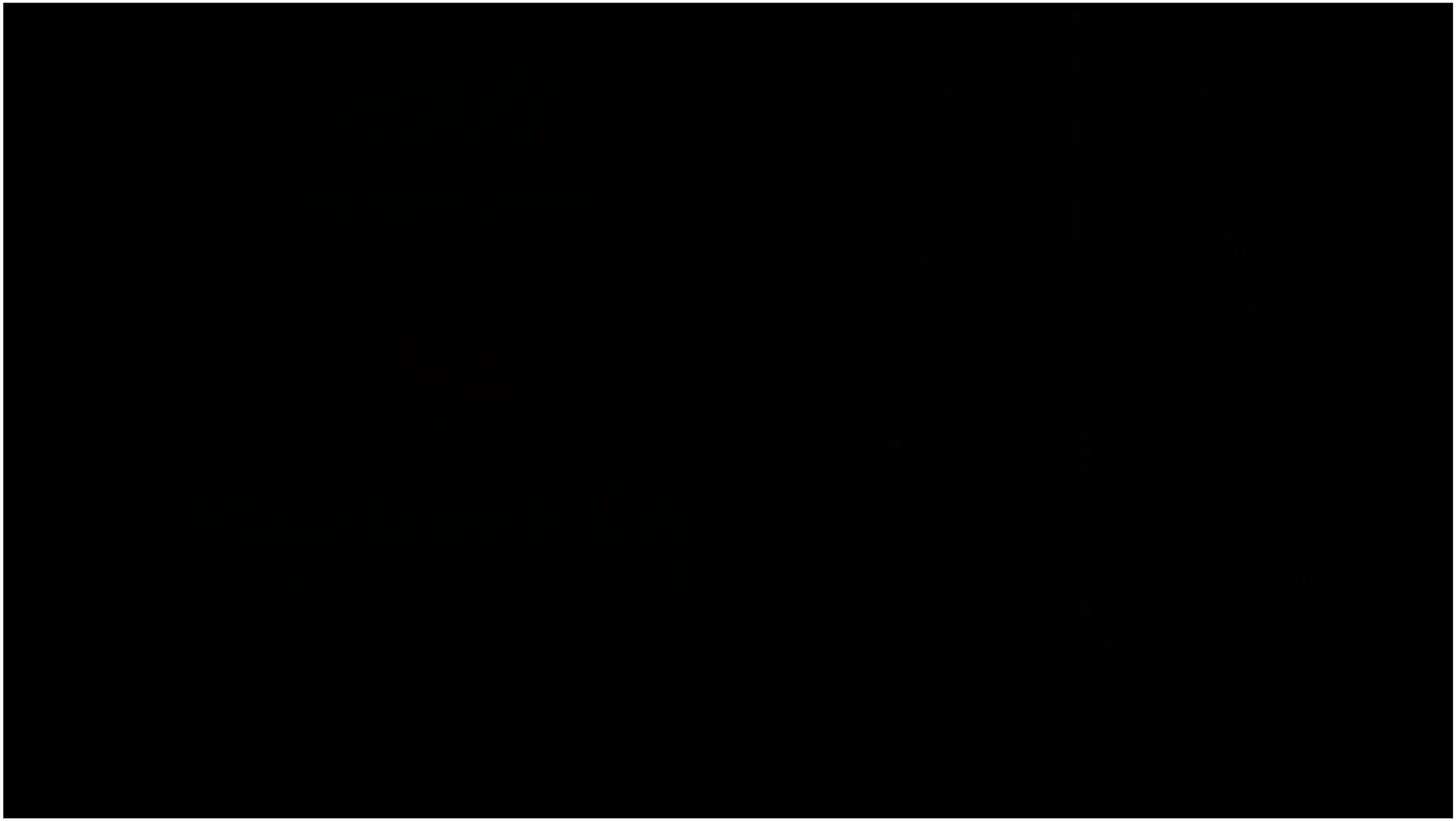
Citable URI: http://hdl.handle.net/1721.1/6125

Date Issued: 1966-07-01

Abstract:

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which allow individuals to work independently and yet participate in the construction of a system complex enough to be real landmark in the development of "pattern recognition". The basic structure is fixed for the first phase of work extending to some point in July. Everyone is invited to contribute to the discussion of the second phase. Sussman is coordinator of "Vision Project" meetings and should be consulted by anyone who wishes to participate. The primary goal of the project is to construct a system of programs which will divide a vidisector picture into regions such as likely objects, likely background areas and chaos. We shall call this part of its operation FIGURE-GROUND analysis. It will be impossible to do this without considerable analysis of shape and surface properties, so FIGURE-GROUND analysis is really inseparable in practice from the second goal which is REGION DESCRIPTION. The final goal is OBJECT IDENTIFICATION which will actually name objects by matching them with a vocabulary of known objects.

2010+: The Golden Age of Computer Vision



http://www.mobileye.com/

2010+: The Golden Age of Computer Vision

http://www.orcam.com/

2010+: The Golden Age of Computer Vision

Amazon's Fire Smartphone Uses "Firefly" Image Recognition

June 18, 2014, 11:28 AM PDT

By Dawn Chmielewski

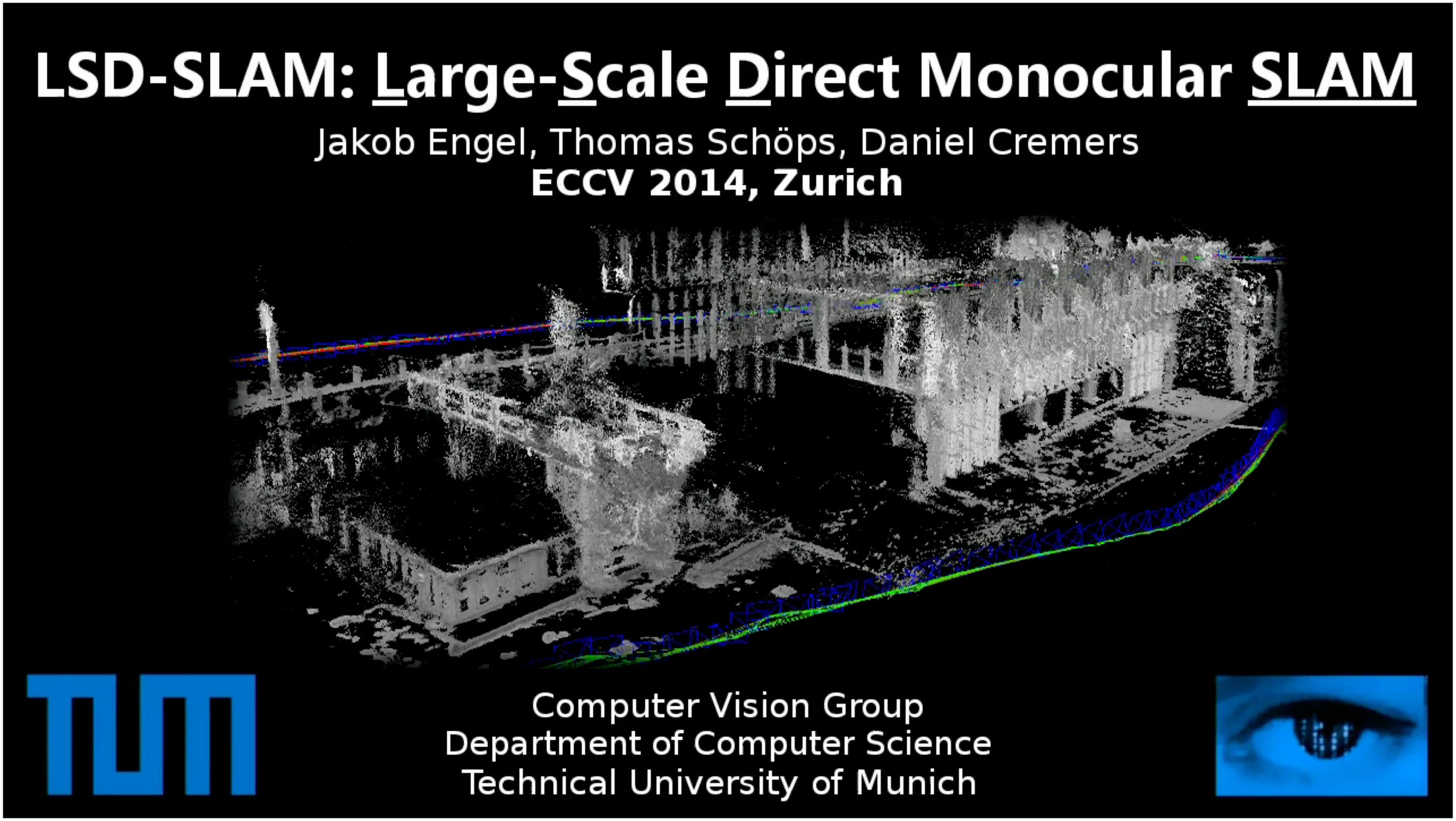
ETHICS BIO

ARTICLES

Amazon's new Fire smartphone will feature advanced image recognition that can use a photo to access media, or look up information about items in the physical world.

This feature is among the ways Amazon is seeking to differentiate its late-to-the-game smartphone from a sea of others in a world dominated by Apple, Google and Samsung. Research firm IDC projected that some 1.2 billion new devices will be shipped this year.

2010+: The Golden Age of Computer Vision



2010+: The Golden Age of Computer Vision

GEORGE PENDLE | JANUARY 5TH 2016

It has been estimated that in 2015 more than a trillion photos were snapped. Instagram, the popular photo-sharing app, claims that some 80m new photos are uploaded to its site each day. We are clearly taking more photographs than ever before, so it's a bit of a

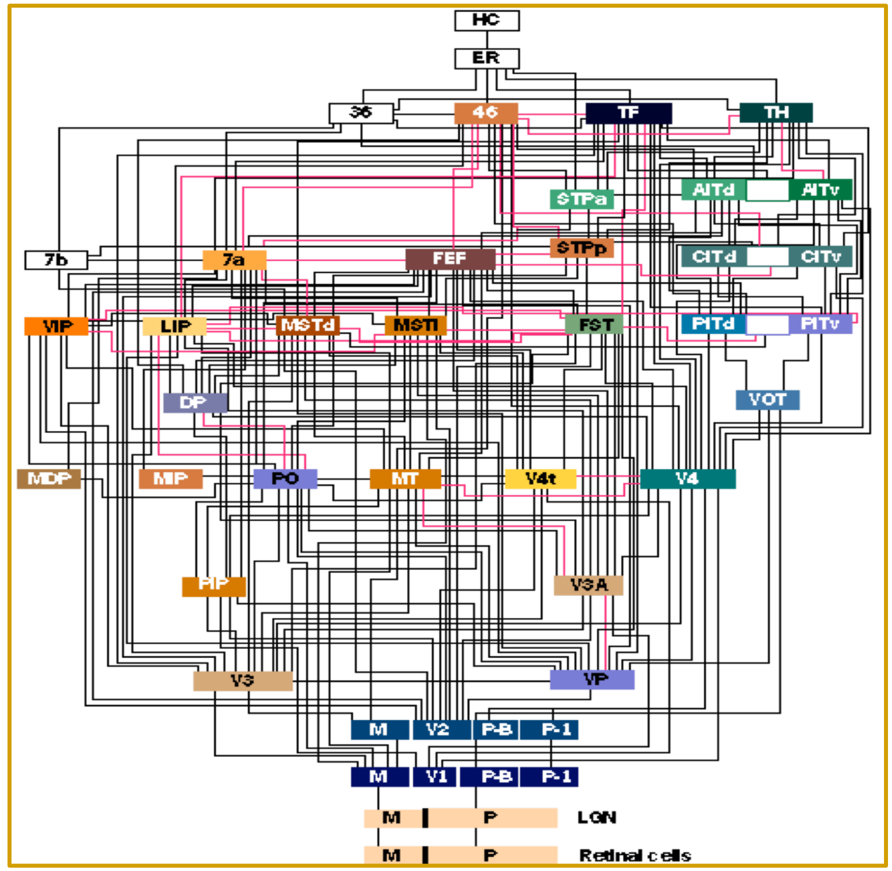
Source: Intelligent Life, The Economist

300m photos uploaded per day on facebook
(Source: Gizmodo)

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:

- 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)

We will cover basic models for understanding ...

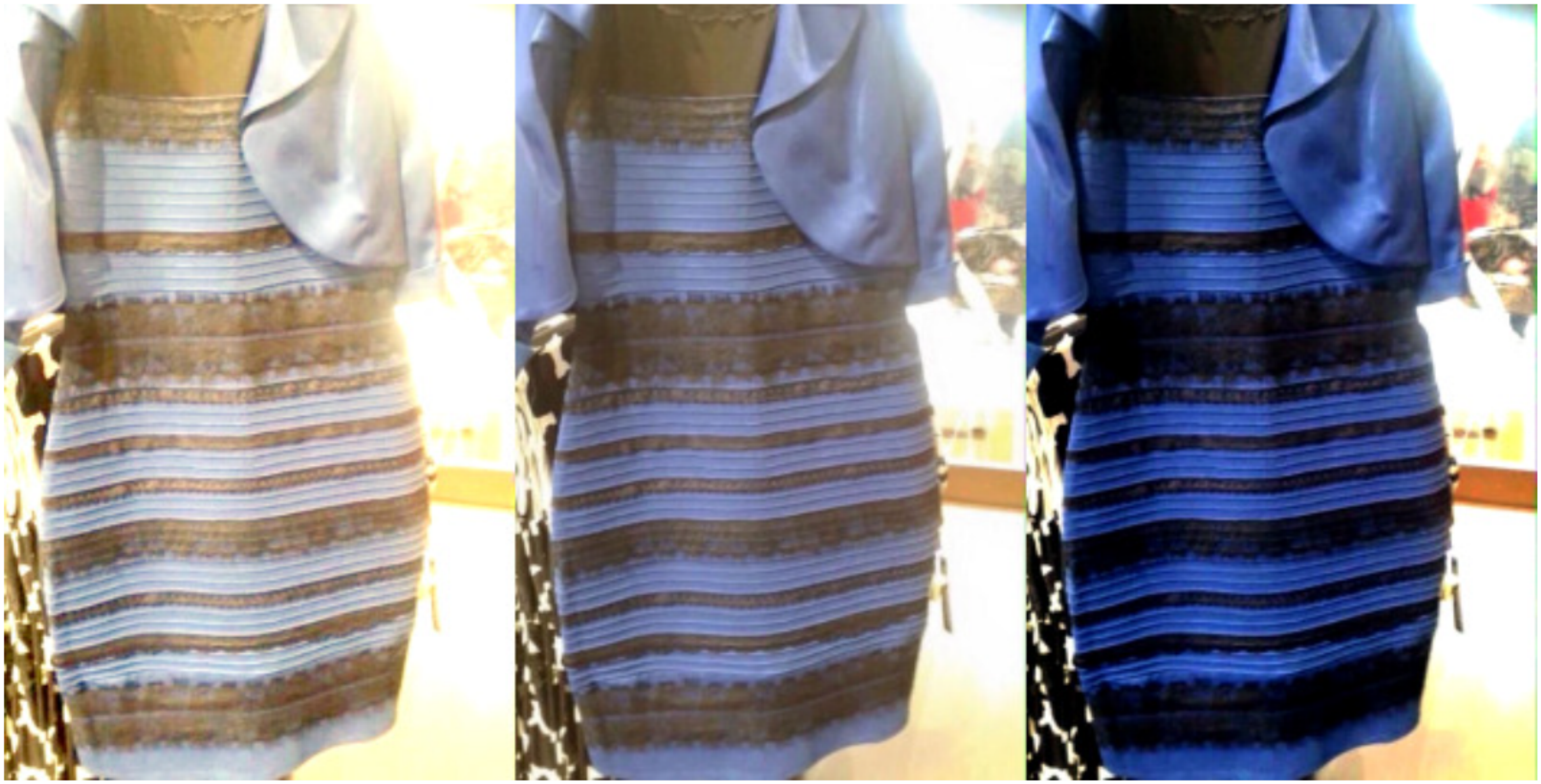
- Lighting and appearance
- Image formation and noise
- 2D motion
- Projection
- How to infer 3D shape from images
- How to track objects in video
- How to represent and recognize 2D and 3D scenes

... through use of key mathematical tools

- Radiometry and reflectance models
- Filtering, sampling and Fourier transforms
- Linear subspace modeling
- Robust estimation
- Bayesian inference
- 3D geometry (Euclidean, affine, projective)

ADAM ROGERS SCIENCE 02.26.15 10:28 PM

THE SCIENCE OF WHY NO ONE AGREES ON THE COLOR OF THIS DRESS



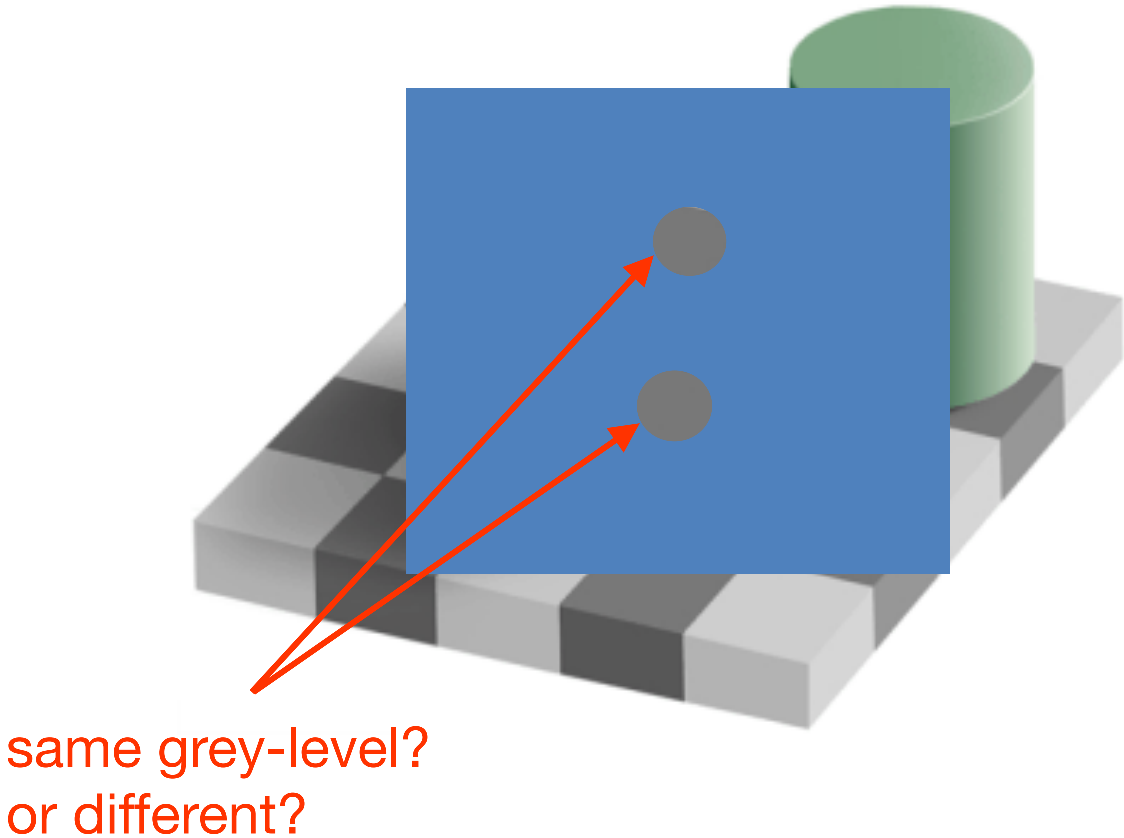
The original image is in the middle. At left, white-balanced as if the dress is white-gold. At right, white-balanced to blue-black. swiked

SHARE

NOT SINCE MONICA Lewinsky was a White House intern has one blue dress been the source of so much consternation.

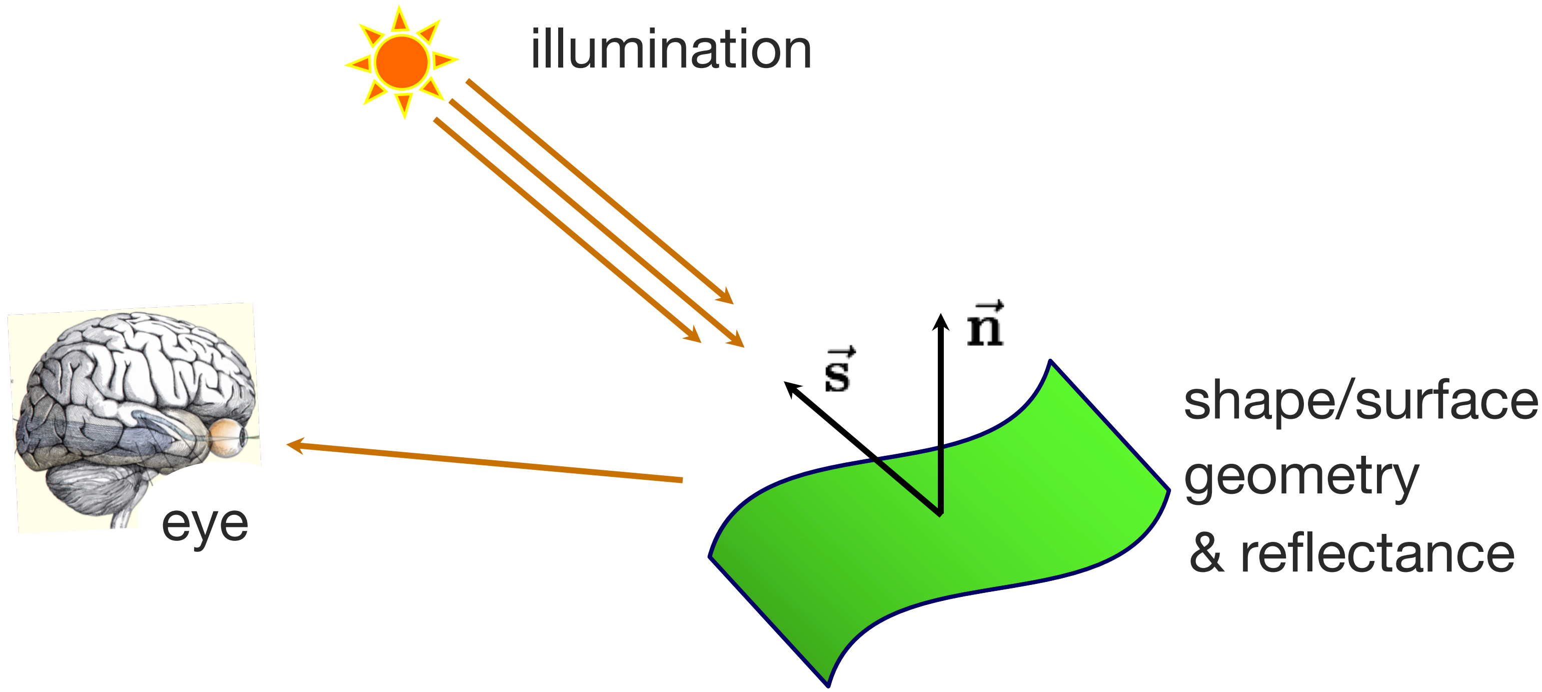
(And yes, it's blue.)

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

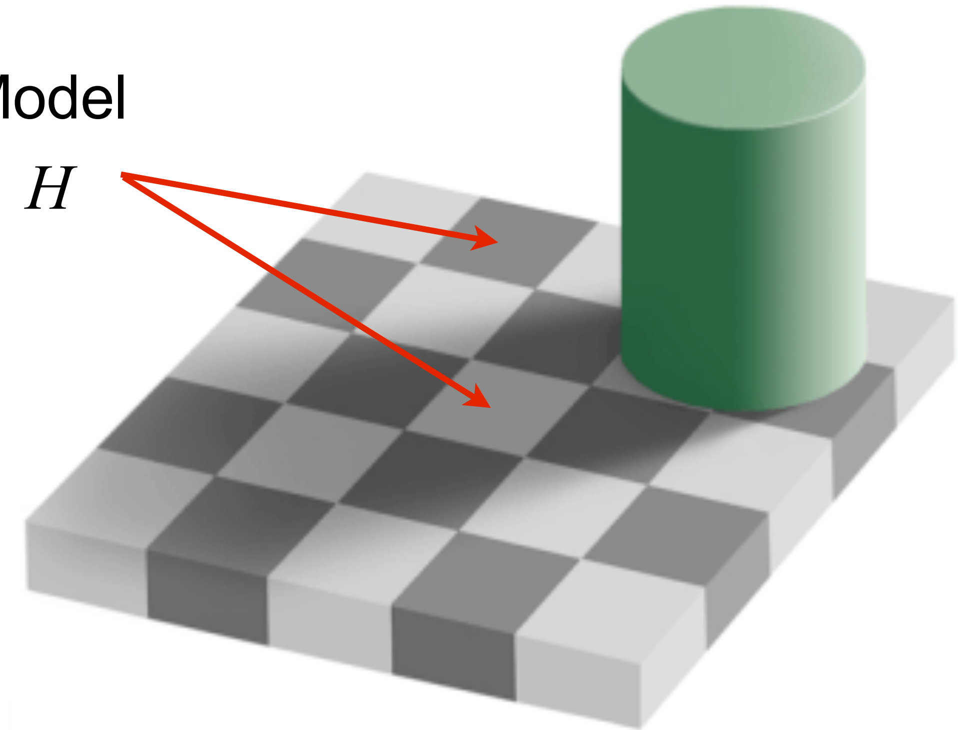
surface normal

light source direction & intensity

Lighting and appearance

Lambertian Model

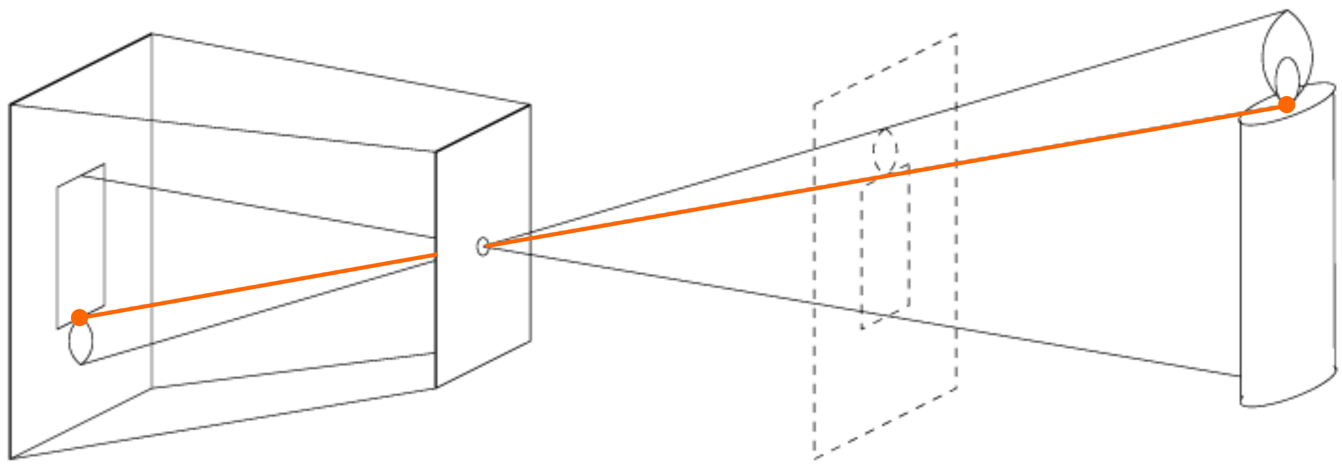
$$L = R \times H$$



Edward H Adelson

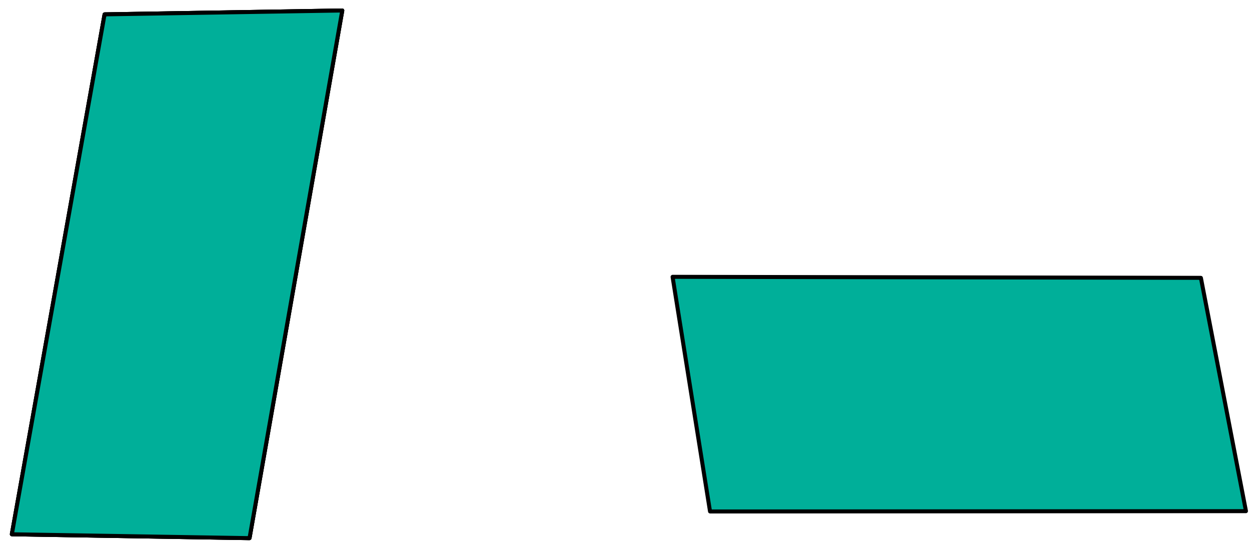
Lighting and appearance

Perspective projection



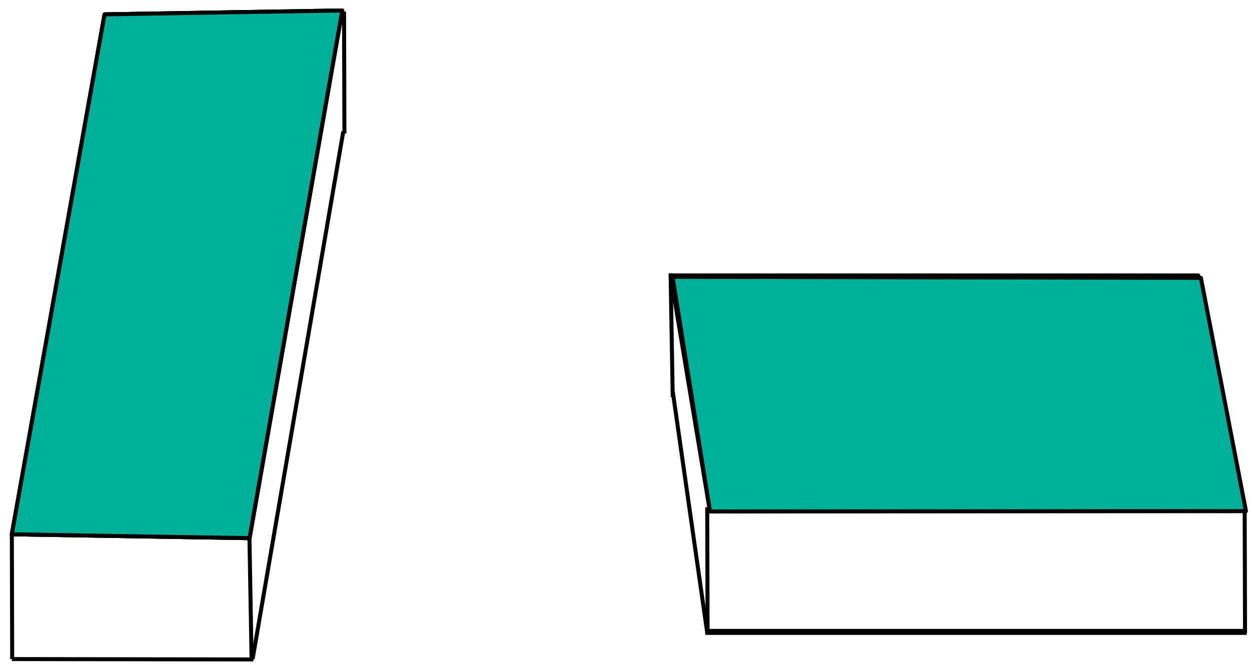
- The projection from 3D points onto the 2D image surface is modeled by perspective projection.
- Depth and size is lost in projection.

Objects are expected to be three-dimensional



Roger Shepard

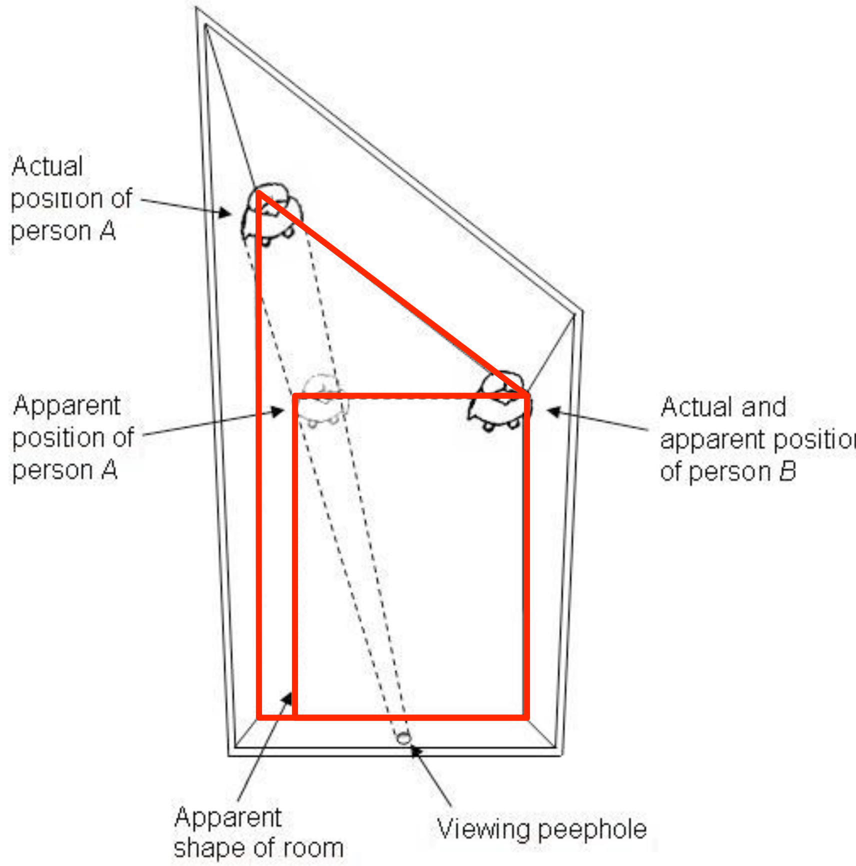
Objects are expected to be three-dimensional



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

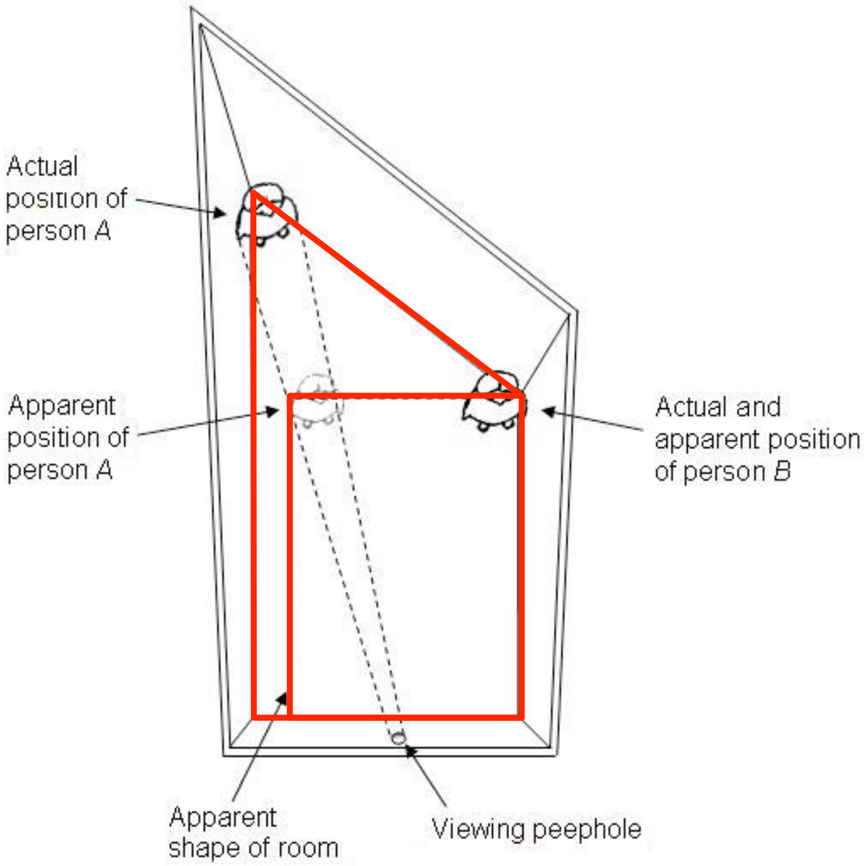
Roger Shepard

Three-dimensional scene understanding



(Ames Room, Adelbert Ames, 1946)

Three-dimensional scene understanding



(Ames Room, Adelbert Ames, 1946)

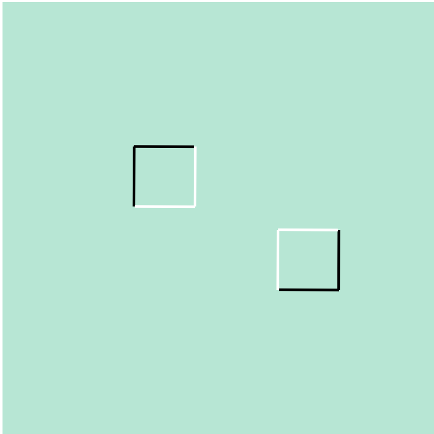
Local consistency and generic views

Jerry Andrus

Illumination is assumed to come from above

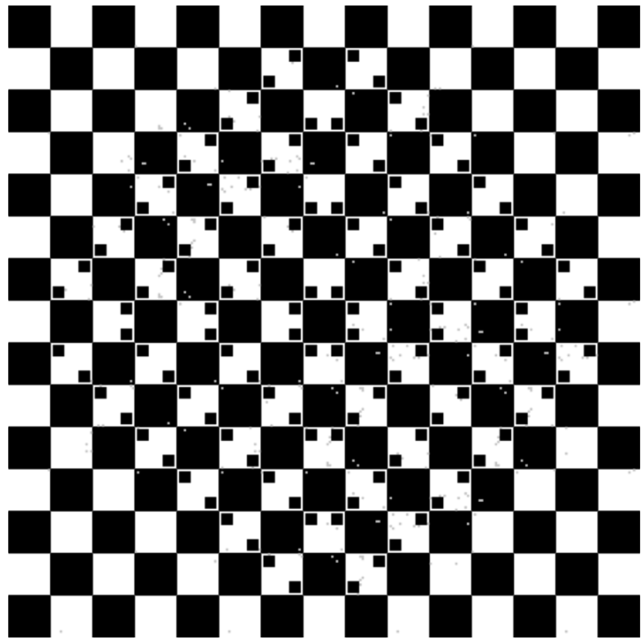
Walter Wick "Optical Tricks"

Illumination is assumed to come from above



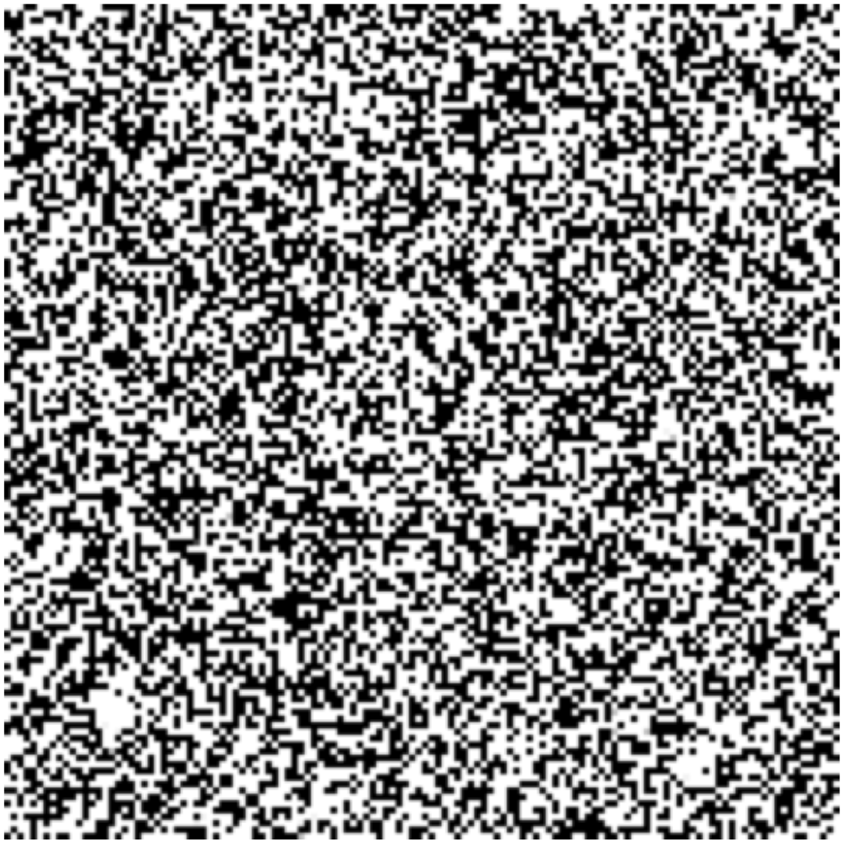
Niko Troje

Perceptual distortion



Akiyoshi Kitaoka

Motion boundaries



Behaviour and intentions

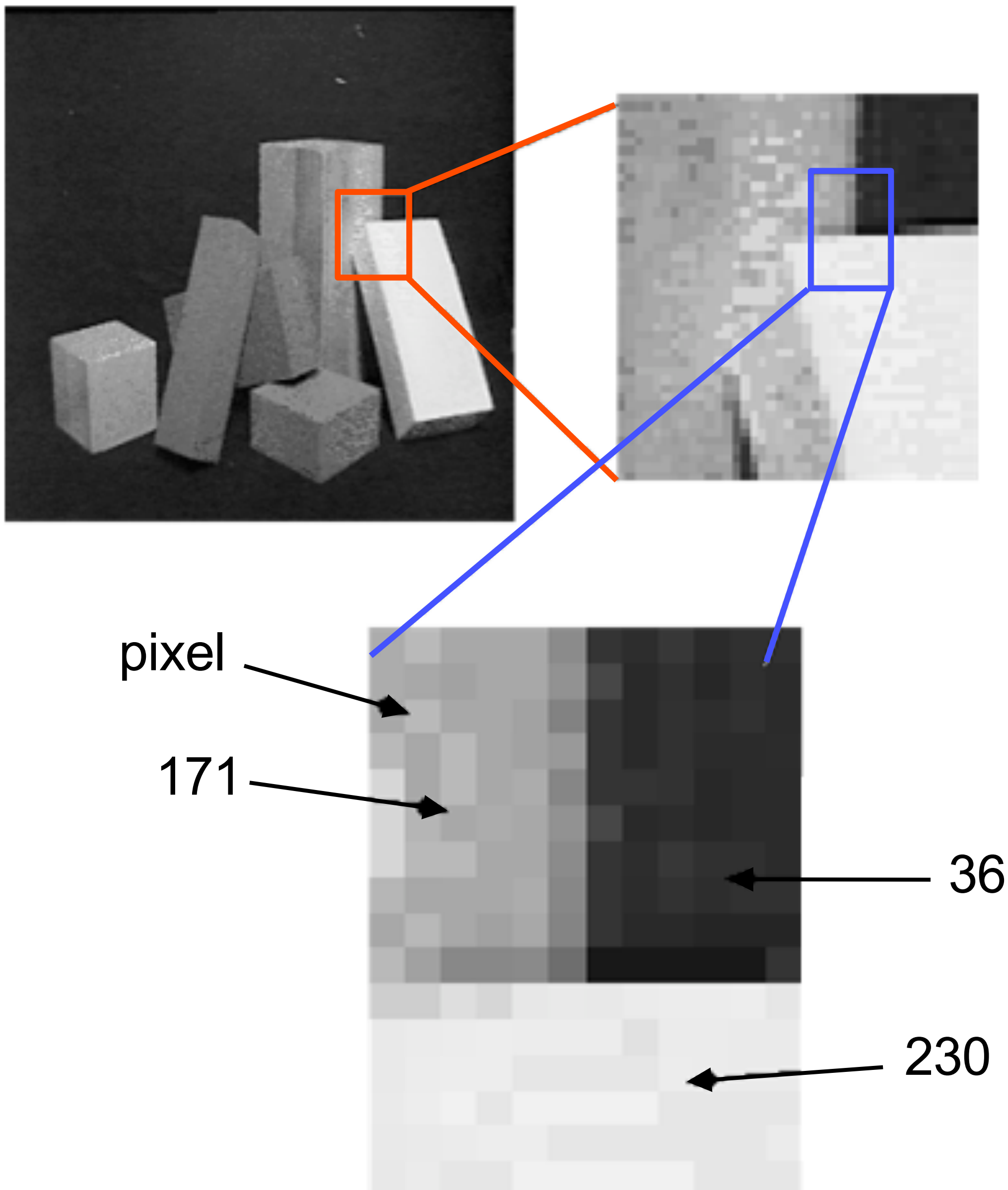
[Heider and Simmel '44]

Computational perception

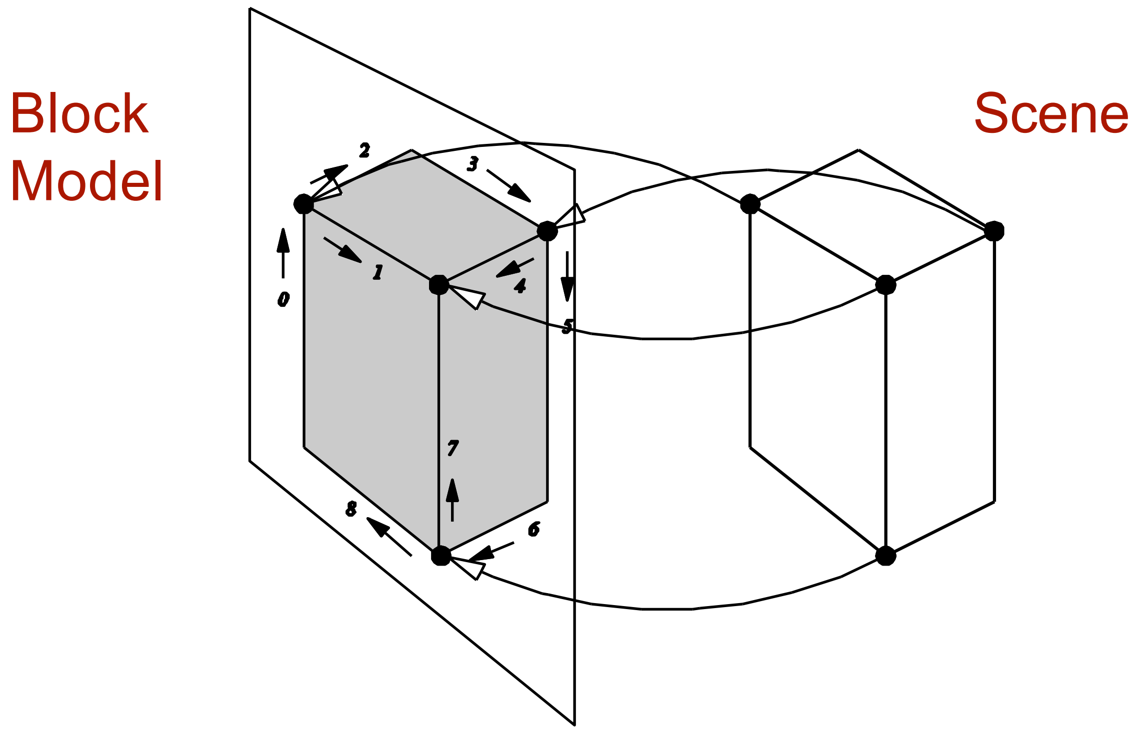
Key elements of computational perception (models):

- 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



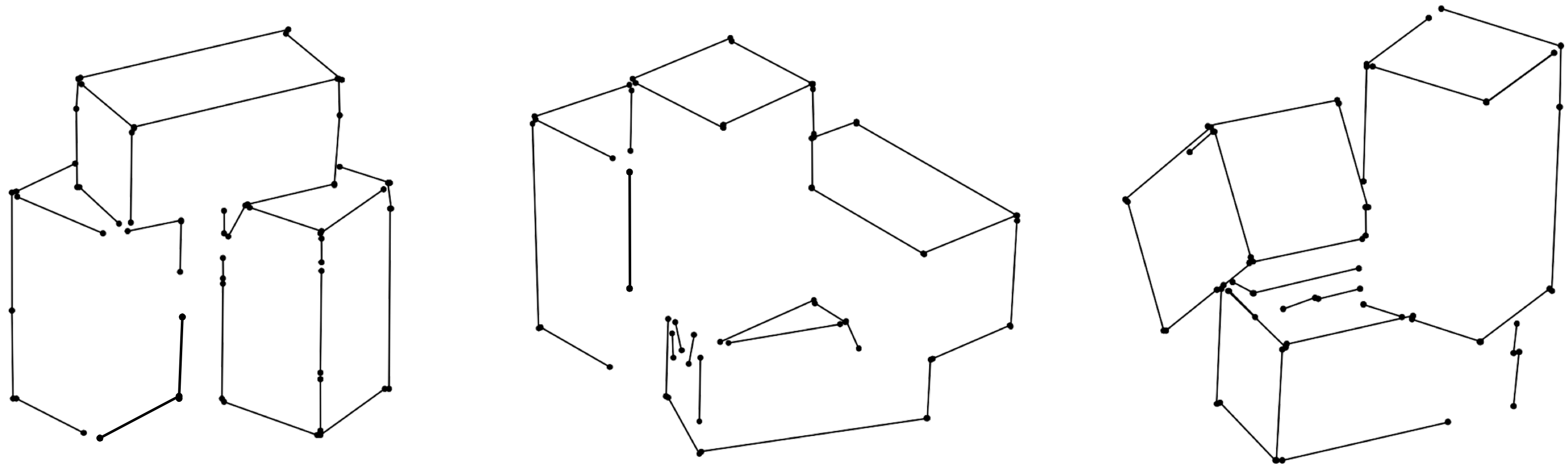
Blocks world domain theory



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

Blocks are opaque, so they can occlude other objects.

Edge measurements



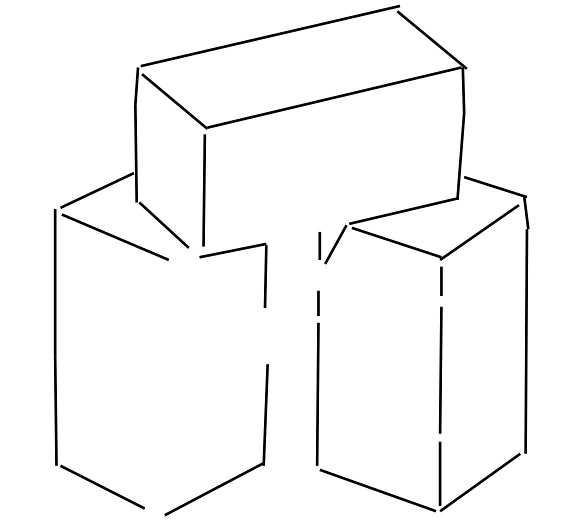
Edges convey useful information about blocks world scenes, but measurements are noisy due to photon noise and modeling error.

Edges are sometimes missing, or broken with missing fragments.

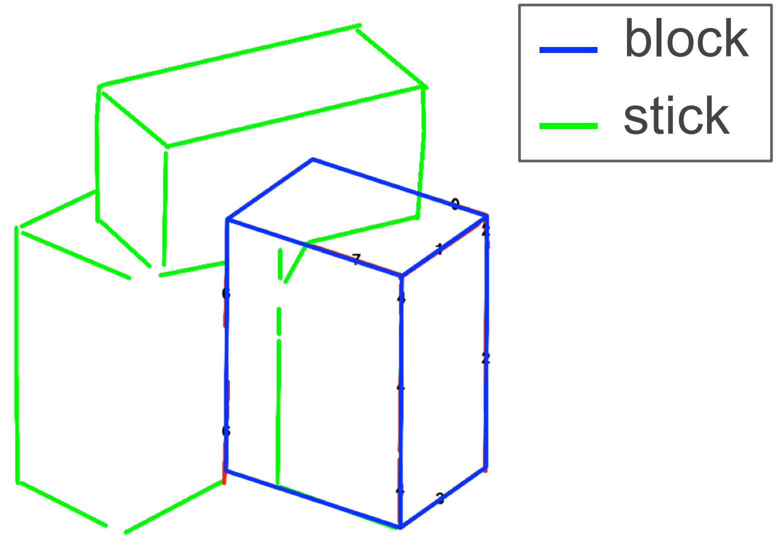
Image interpretations explain all edges in terms of blocks & sticks.

Slides by D. Fleet and A. Jepson

Consistent interpretations

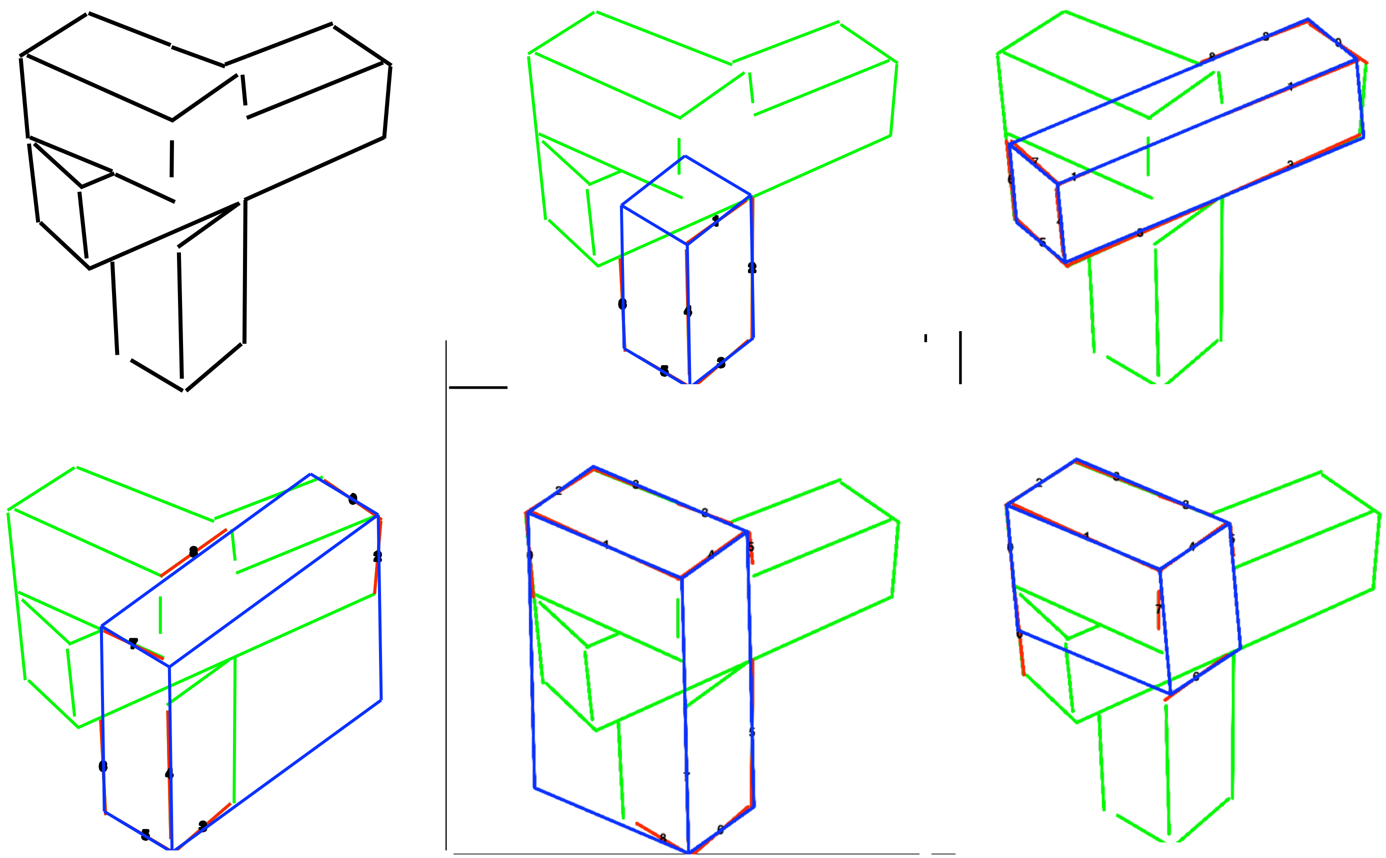


Edge segments
(34 edges here)



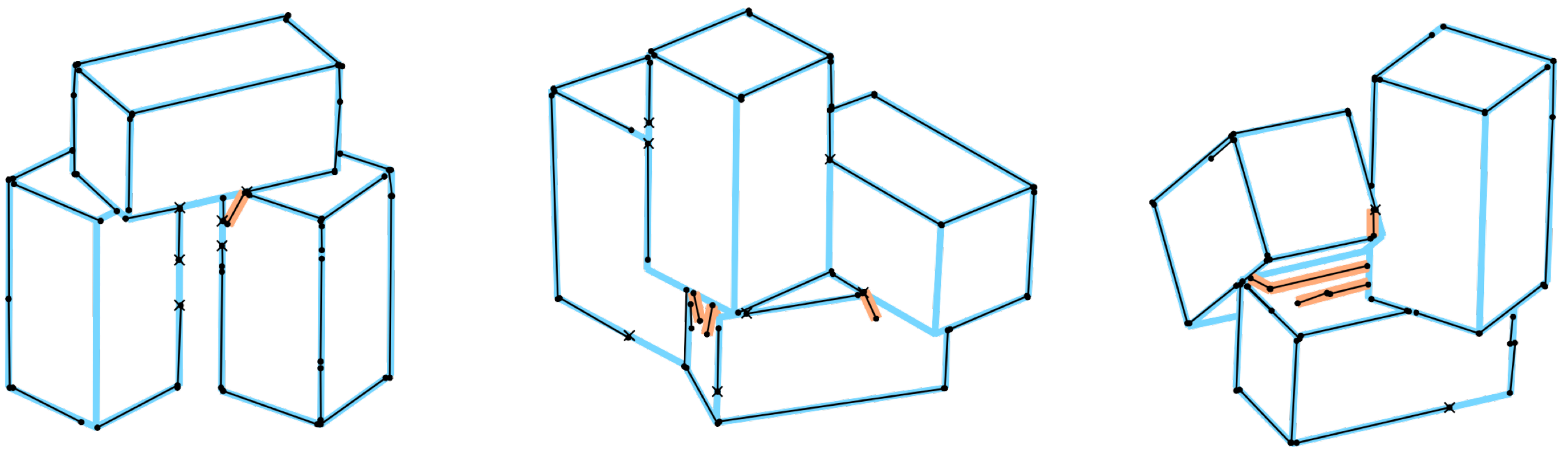
Candidate block
(interpretation)

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

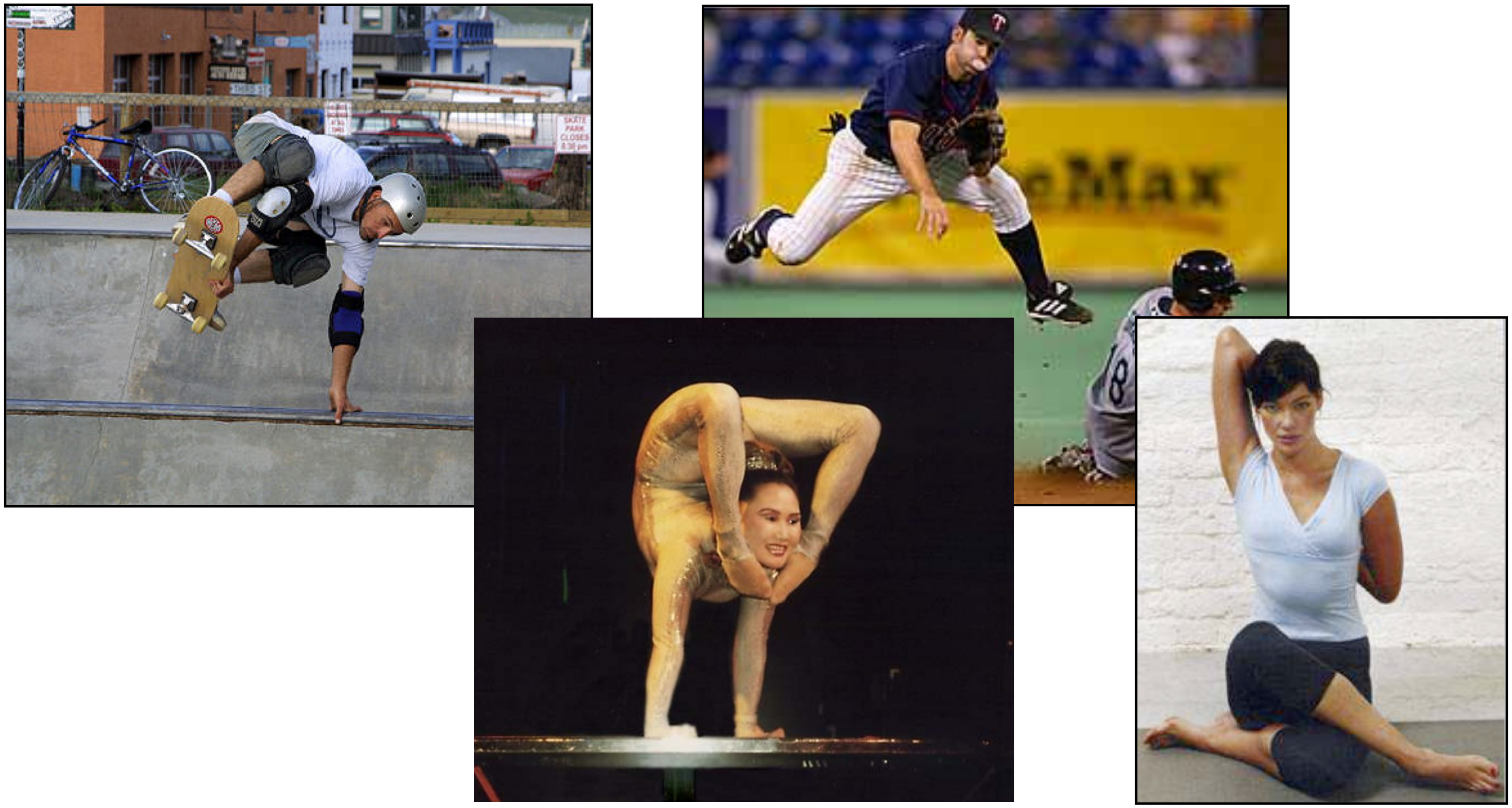
Slides by D. Fleet and A. Jepson

Challenges: Appearance, size and shape

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

Slides by D. Fleet and A. Jepson

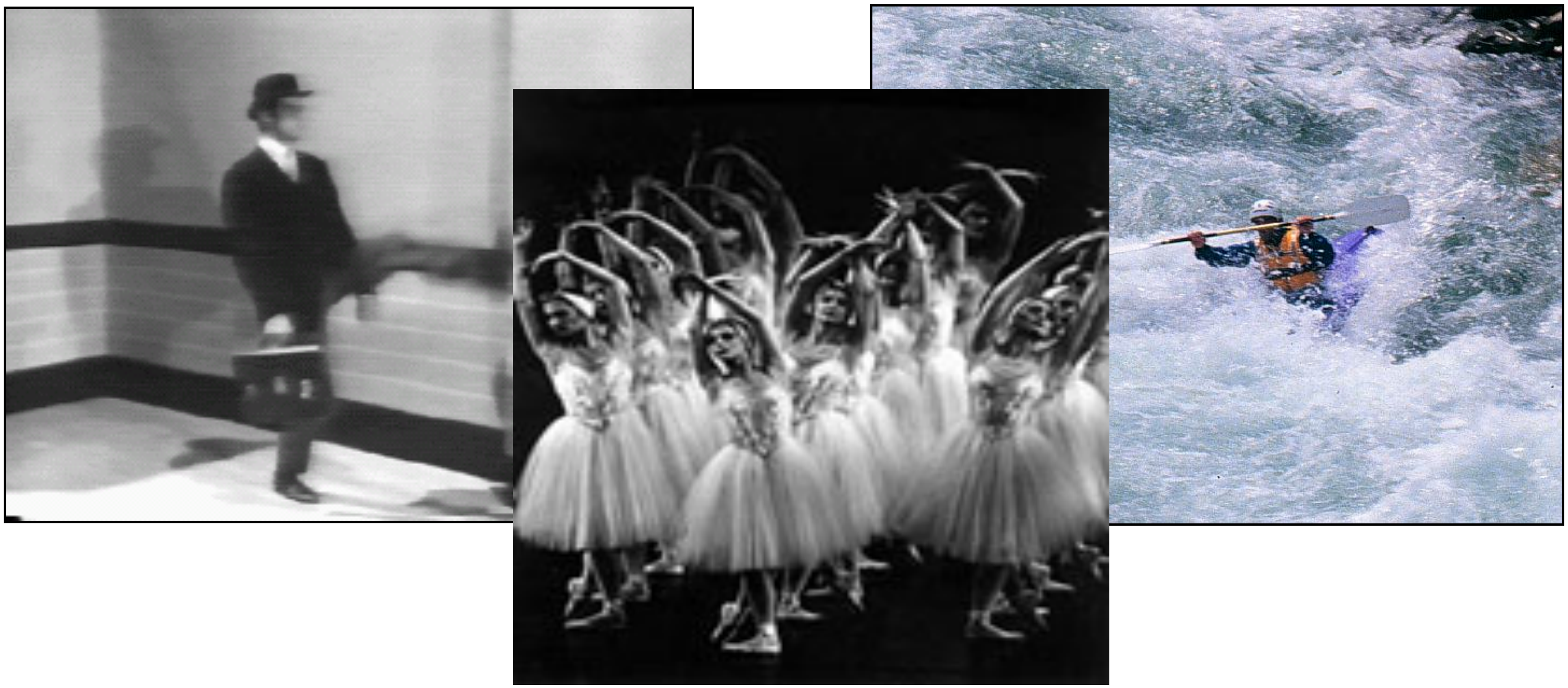
Challenges: Complex pose / motion



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

Slides by D. Fleet and A. Jepson

Challenges: Noisy and missing measurements



Ambiguities in pose are commonplace, due to

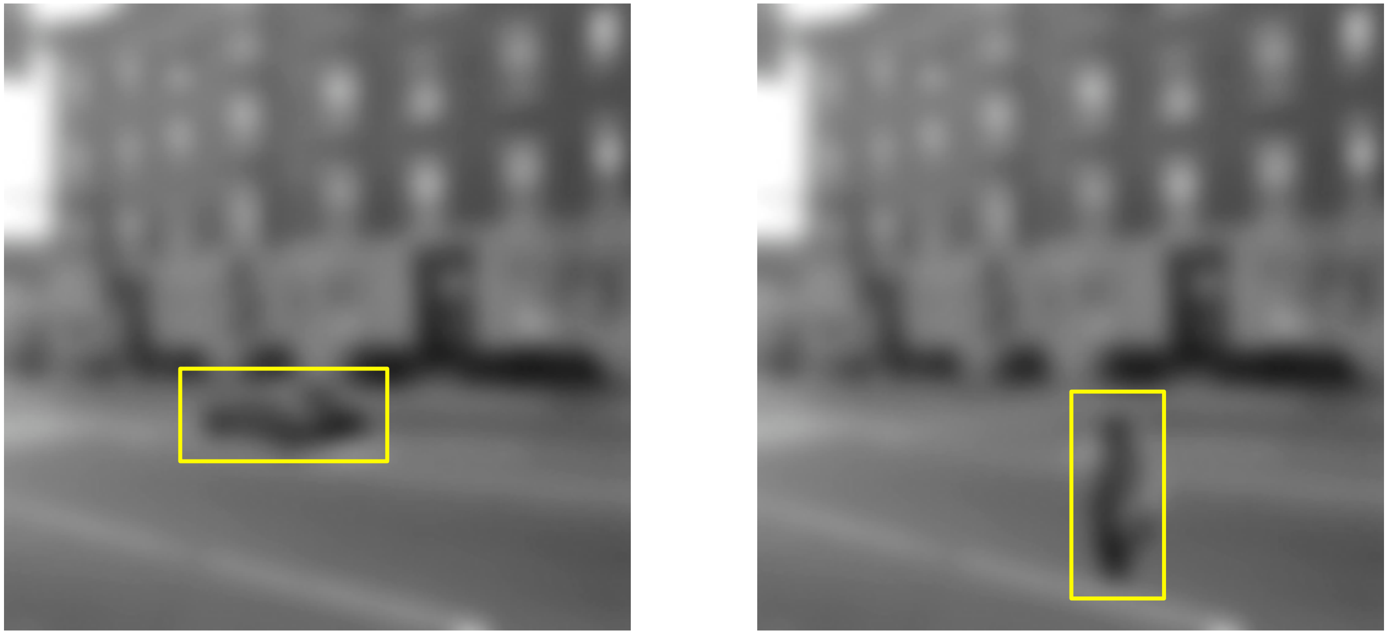
- background clutter
- apparent similarity of parts
- occlusions
- loose clothing
- …

Slides by D. Fleet and A. Jepson

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]

CSC 2503: Basic aims

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?

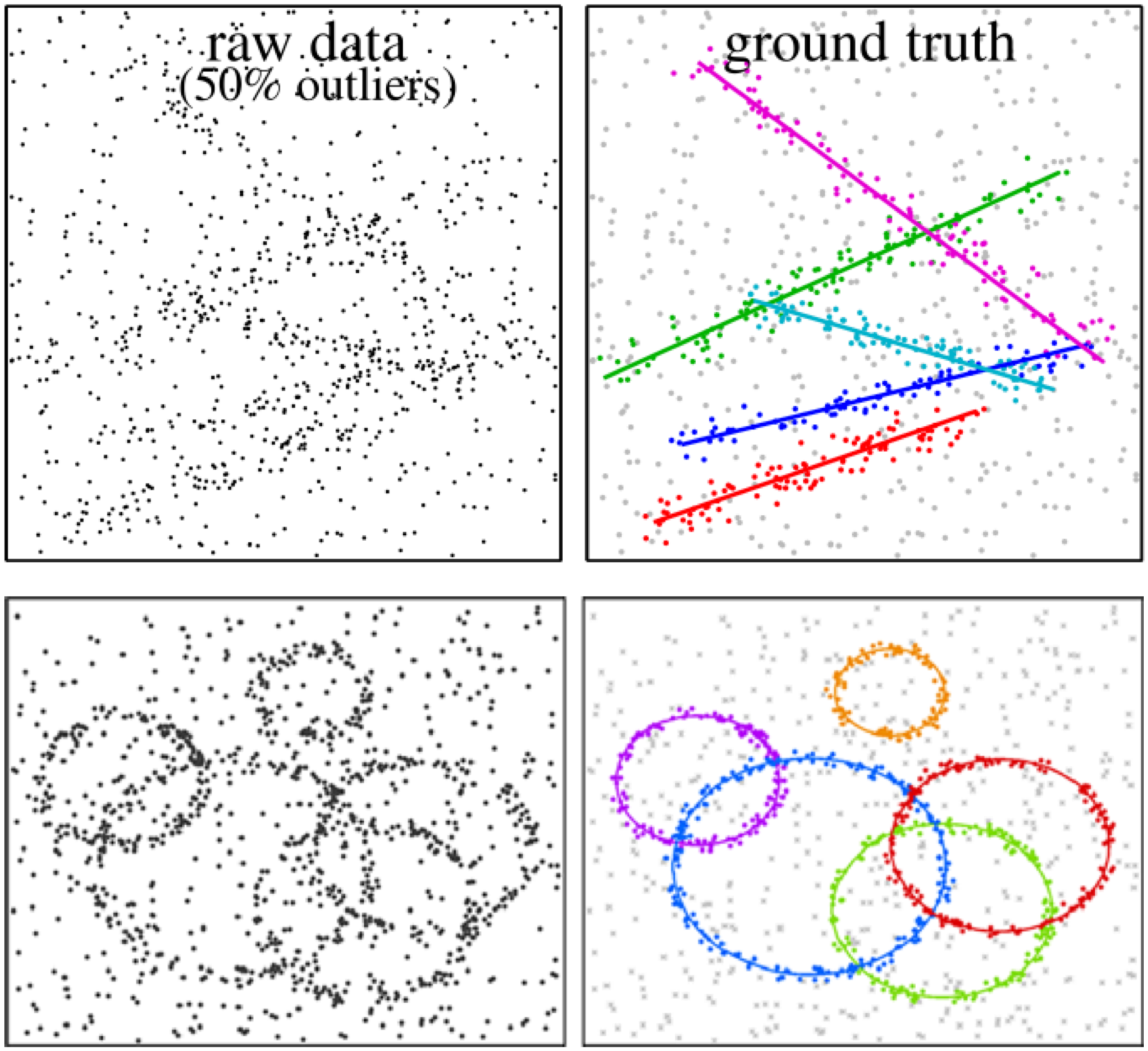
Current practice – simple models:

- 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.

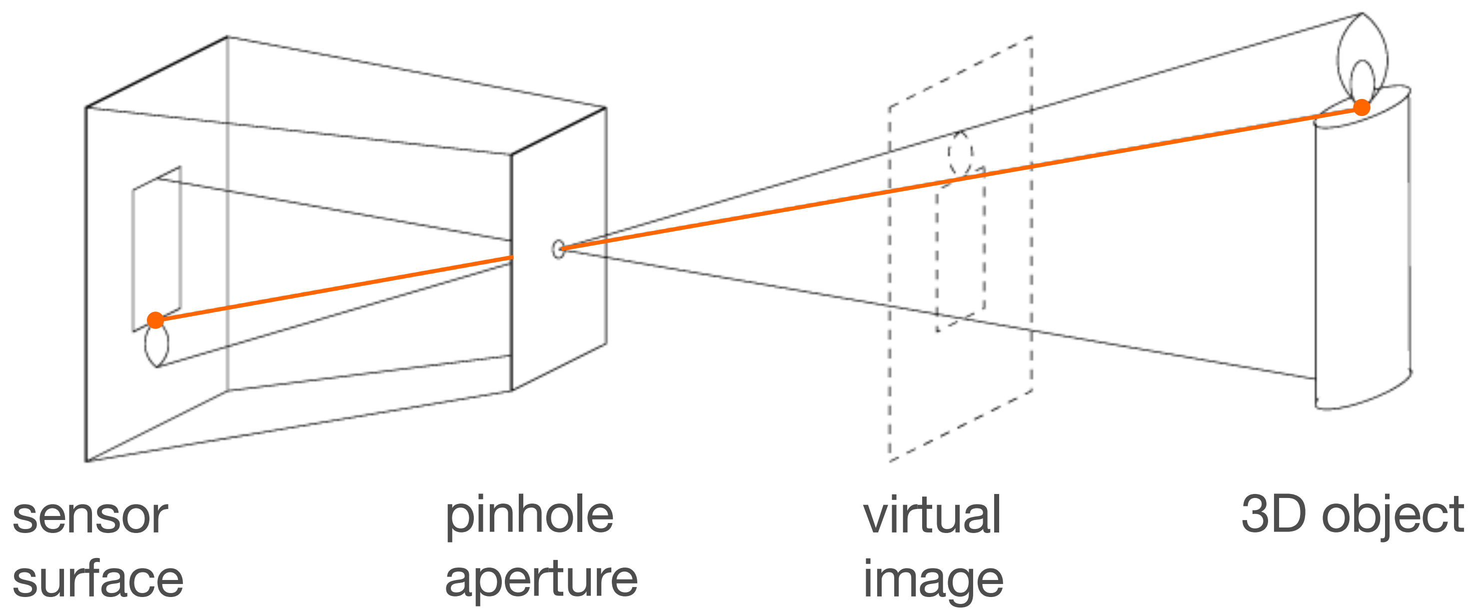
Course outline

- Lecture 1: Model fitting & robust estimation



Course outline

- Lecture 1: Model fitting & robust estimation
- Lectures 2 & 3: Geometric Computer Vision



Course outline

- Lecture 1: Model fitting & robust estimation
- Lectures 2 & 3: Geometric Computer Vision

Reconstructing the World* in Six Days
***(As Captured by the Yahoo 100 Million Image Dataset)**

Jared Heinly, Johannes L. Schönberger,
Enrique Dunn, Jan-Michael Frahm

The University of North Carolina at Chapel Hill

CVPR 2015

Course outline

- Lecture 1: Model fitting & robust estimation
- Lectures 2 & 3: Geometric Computer Vision
- Lecture 4: Image filtering & Fourier transforms

Course outline

- Lecture 1: Model fitting & robust estimation
- Lectures 2 & 3: Geometric computer vision
- Lecture 4: Image filtering & Fourier transforms
- Lecture 5 & 6: Fundamentals of image formation

Credits: dpreview.com

Course outline

- Lecture 1: Model fitting & robust estimation
- Lectures 2 & 3: Geometric computer vision
- Lecture 4: Image filtering & Fourier transforms
- Lecture 5 & 6: Fundamentals of image formation
- Lecture 7: Probabilistic appearance models
- Lectures 7 & 8: Analyzing image motion
- Lecture 9 & 10: Advanced inference (deep nets, MRFs)

Making the most out of the course: lectures

Before each lecture:

- Read & try to follow last year's lecture slides (already on dropbox)
- (optional) Make a list of questions, to be asked/clarified in class
- Print the lecture slides (or put on a tablet) and bring them to class

During each lecture:

- Take notes (eg. on the printed slides)
- If I'm going too fast, say so!
- ASK QUESTIONS: No question is too dumb
- Don't assume others can follow the lecture better than you: if you have a question, you may do many others a favor by asking it
- Remember the class has students from many backgrounds (CS, Engineering, Grad, Undergrad, non-UofT)

What to expect in each lecture

- Goal is to give intuitions & basic mathematical tools... so that YOU can work out the parts I skip
- Parts of each lecture will go into a lot more detail than others

Making the most out of the course: lectures

After each lecture:

- Make sure you can derive all EQs/slides marked with a *
- If you can't, ask your fellow students (see below), ask your questions in the tutorial, or come to my office hours

Making the most out of the course: study groups

- You will get the most out of the course by forming a 2-3 student informal group (for bouncing questions, understanding the assignments, following the readings). You can use Piazza to do this
- You can also use Piazza to post questions to your fellow students

Making the most out of the course: readings

Syllabus

- Goal is NOT to provide broad survey of techniques; it is to cover key foundational principles (that have stood the test of time)
- Course covers a lot of material, only partly discussed in lectures and tutorials
- A good understanding of ALL lecture slides & notes is required; if you cannot follow something, discuss in your study group, post on Piazza, talk to me, or ask in tutorial
- The remaining readings (book chapters, papers, etc) are not required (but will deepen your knowledge)
- By end of course you should be able to understand and appreciate research papers from recent vision conferences (regardless of their specific topic)

Past exam questions: don't wait until the end of the course to look at them. Tackle them throughout the course to assess your understanding of course content!

Making the most out of the course: tutorials

- Come prepared: read the slides & notes from week before, have your questions ready
- TA will be instructed to prioritize your questions, and only then go over selected portions of the lecture slides
- First tutorial: next week

Making the most out of the course: assignments

- Can only cover a relatively small fraction of the course content
- Read & understand the assignments early, bring your questions to tutorial & office hours
- A1 will be on dropbox next Wednesday (due 3 weeks later)
- 3 assignments in total

Academic Honesty

- Assignments are strictly individual work
- Your assignment work (programming AND theoretical) should not be discussed with anyone. This includes fellow students, students who have taken the course in the past, etc.
- You should not ask others assignment-related questions that you would expect I or the TA would not answer
- Searching for solutions via electronic means (web, repositories, etc) is strictly forbidden

- I have a zero-tolerance policy on academic honesty (and enforce it rigorously)

CSC 2503: Logistics

Tutorials:

- Fridays 11-noon
- First tutorial this Friday
- Tutorial classroom: GB 119
- TA: Parsa Mirdehghan (plus one more)

Assignment:

- First assignment out next week, due in 3 weeks