CS 699i, Spring 2005, part III

Computer Vision

(Making Machines See)

Kobus Barnard

Computer Science, University of Arizona

“ …, vision is the process of discovering from images what is present in the world, and where it is.

… our brains must be capable of representing this information … ”

Marr 82, page 3.

Vision Systems

Biological

eye + brain

Man made

camera + computer

Data ---> Efficacy

Vision and Ambiguity

Given visual stimulus (image), I, there are many possible realities or worlds, W_i, that can explain it

Hence the process going from

I --> W

is inherently ambiguous and difficult

Vision versus Graphics

Graphics

model of the world --> image

Vision

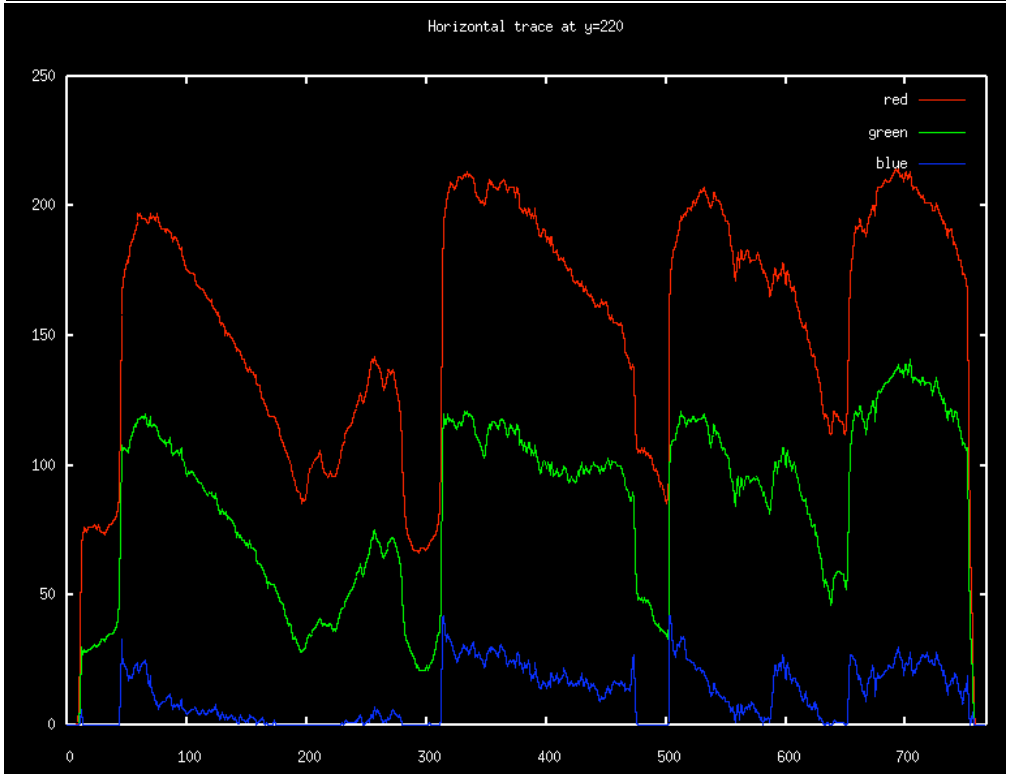
image --> model of the world

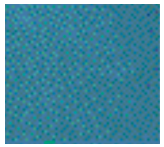
(which one?)

(need to be smart!)

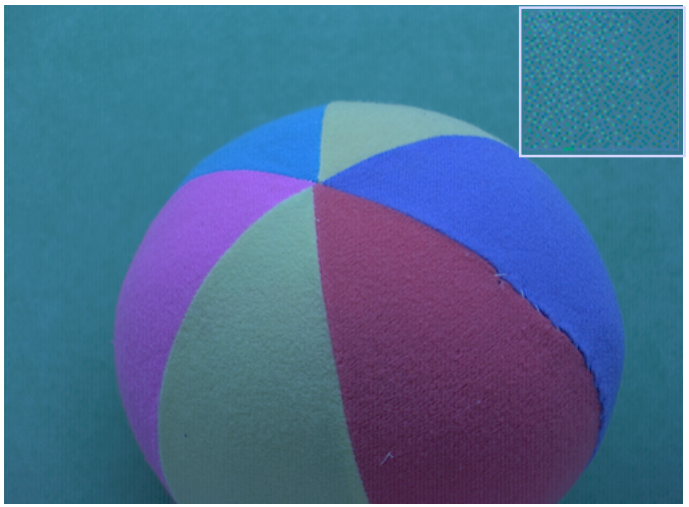
This could be a picture of water or it could be water. The two visual stimulus are similar, but what is in the world is quite different! Further the assumption that the waves are roughly the same size yields inferences about depth.

Shading Cues





Color signal ambiguity. This could be a blue patch under white light, or a white patch under blue light.



(Same scene, but different illuminant)

Vision as Computation

Visual Representation ⇒ Semantic Representation

⇒ A tiger lying in the grass

Vision as Computation

Can the plan really work?

Two possible problems

Not every conceivable set of input and output is linked by an algorithm (there are things that computers cannot compute, regardless of how fast they become!).

Even if vision is computing, there may not exist an algorithm which can run fast enough on conventional hardware (NP complete or worse).

Vision as Computation

If computers were infinitely fast, would that solve the problem?

A dumb algorithm*:

Generate all reasonable worlds (to some tolerance)
Produce all images that your camera could take
Compare results with the image being analyzed

*This does NOT answer the posed question, but could be a starting point in thinking about it.

Philosophical Stance

Vision is computational.

If you have a theory about how it works, you should be able to write the code to demonstrate it.

Vision as Statistical Inference

Consider some world, W, with probability p(W)

Consider your observation, O

If you have a statistical model for W, then you can compute p(O|W)

Of course, what you want is p(W|O)

Fortunately, Bayes rule links them

$$p(W|O) \quad P(O|W)p(W)$$

History, etc.

History of the field is largely about failure, and learning what the hard parts are.

Many things have been learned about inferring useful bits of information from images but robustness is lacking

Representation and linking to world knowledge is poorly understood

General object recognition remains a big open problem

Section outline

- Introduction and lessons from human vision
- Introduction to theory of computation
- Bottom up and top down processing.
- Ambiguity (examples: color constancy, inferring depth)
- Vision as statistical inference
- Representation (example: translation models for recognition)
- Task oriented evaluation and applications

Assignment(s)

UNDER CONSTRUCTION

- Will involve programming which implements a bit of math
- Will provide some flavour of modern algorithms
- Will solve a intractable problem with an iterative approximation
- Will have a written component providing your ideas about what programs of this genre can and cannot say about computational intelligence

Kobus Barnard

Gould-Simpson 730

kobus @ cs.arizona.edu

http: // kobus.ca

Office Hours by electronic signup : Th, 3:30 -- 5:00

Login: public

Password: meetkobus

Class materials will be put online, but will be password protected if you are coming from off-campus:

Login: me

Password: vision4fun