

Computational Vision

U. Minn. Psy 5036

Daniel Kersten

Lecture I

The course

Syllabus is online at courses.kersten.org

- Prerequisites
- Readings and lecture material
 - a "lab course"
- Grading

General class goals are to learn...

- about the primate visual system, including your own
- an interdisciplinary approach to a complex problem
- statistical techniques for making good guesses from lots of weak information
- how knowledge of vision can be applied to everyday problems
- a high-level programming language
- to improve your writing

Introduction to the problem of vision

Understanding visual perception is an important problem in psychology

One of the great mysteries of psychology is how the human visual system determines what and where objects are just by looking. This is one definition of the problem of vision (Marr, 1982). The perception of what is out there in the world is accomplished continually, instantaneously and usually without conscious thought. The very effortlessness of perception disguises the underlying difficulty of the problem.

Vision is important because it is one of the principle routes to our acquisition of knowledge, as well as guide to its utilization. It takes just one quick glance at the picture in the Figure below to recognize the fox, a tree trunk, some grass and background twigs. But that is just the beginning of what vision enables us to do with this picture. With a few more glances, one can see, as one person describes it, that:

“It is a baby fox emerging from behind the base of a tree not too far from the viewer. It is heading right, high-stepping through short grass, and probably moving rather quickly. Its body fur is fluffy, reddish-brown, relatively light in color, but with some variation. It has darker colored front legs and a dark patch above the mouth. Most of the body hairs flow from front to back...and what a cute smile, like a dolphin.”

The ability to generate an unbounded set of descriptions from a virtually limitless number of images illustrates the extraordinary versatility of human visual processing

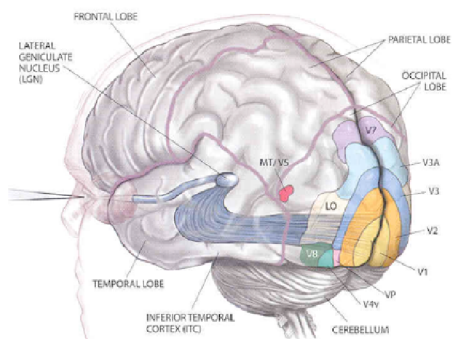


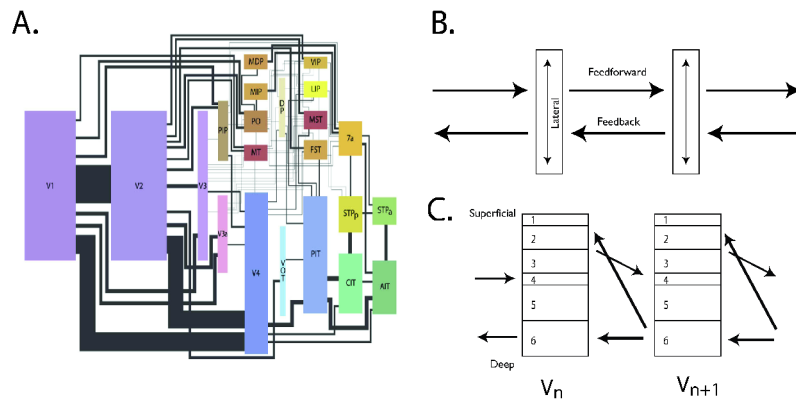
Understanding vision is an important aspect of brain science

With about 10 million retinal receptors, the human retina makes on the order of 10 to 100 million light measurements per second. These measurements are processed by about a billion plus cortical neurons.

Visual neuroscientists currently estimate that 40 to 50% of human visual cortex (your gray-matter) is closely involved in visual processing. The general structure of cortical layers and pattern of inter-connectivity is similar across the neocortex. Thus the hope that if we can understand visual computations in the cortex, this knowledge may generalize to other cognitive domains.

Panel A of the figure below shows a subset of the visual areas of the monkey, together with their interconnections. The areas of the rectangles represent the relative proportions of neurons in each, and the thickness of the lines the proportion of feedforward neural fibers connecting them.





Vision is a complex process whose full understanding requires a computational perspective, including mathematical modeling and programming tools.


Vision is a challenging mathematical and computational problem

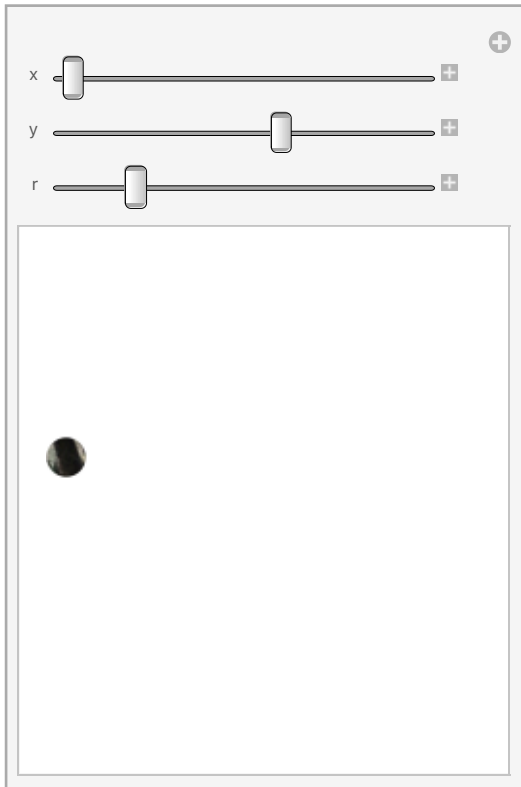
The problem of vision is not only important from the point of view of understanding the brain, but it is also theoretically challenging. Formal solutions have implications for robotics and artificial intelligence. As such vision is an active area of research for computer scientists, engineers, and mathematicians as well.

Why is it a hard formal problem?

A central mathematical problem is the multiplicity of possible scene or object configurations that could have caused a particular set of image measurements. Richard Feynman compared the problem of vision to deciding what jumped into a swimming pool just by measuring the bobbing water height as a function of time using a ruler in the corner of the pool. There are lots of ways of in which a given pattern of water heights could arise. Further, as with water height, image measurements are very indirectly related to useful scene information--was it a beach ball or infant that fell in? An understanding of image formation and optics does not sufficiently constrain the number of possible scene descriptions, \mathbf{S} that could have given rise to any one image, \mathbf{I} . This is one of the defining characteristics of inverse problems in general--the data underconstrain the solution.

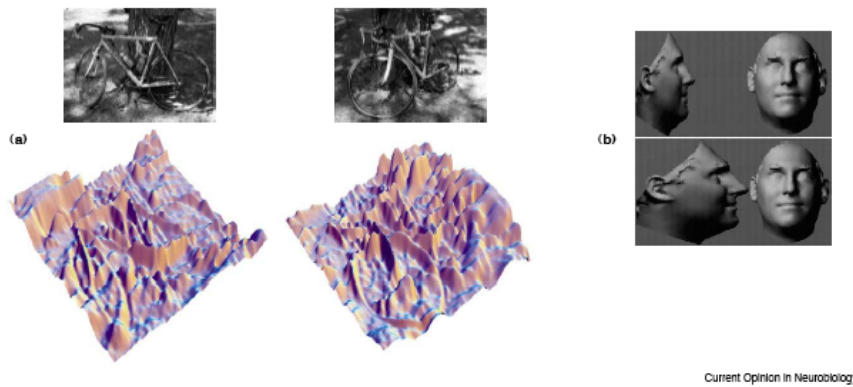
Here is a demonstration of the lack of locally recognizable features in a typical "natural" image -- the problem of *local ambiguity*.

```
Manipulate[
  ImageApply[UnitStep[#] &, , Masking -> Graphics[Disk[Scaled[{x, y}], r]]],
  {x, 0, 1}, {y, 0, 1}, {{r, .1}, .01, .5}]
```



Note that for small apertures, you can describe basic features within a patch, such as the colors, locations and orientations of edges, but you may not be able to tell which bits within the image belong to a surface that is in front or behind another. As the aperture gets bigger you can begin to see surface properties (e.g. fur, or bark) and see surface boundaries, i.e. to tell where one surface ends and another begins, but you may not be able to recognize what object category the pixels are coming from. When the aperture is sufficiently big, then you can tell that a patch is a leg, or a fox, or a tree.

Not only can the same object can give rise to different images (panel a below), but different objects can give rise to the same image (panel b below).



Panel b illustrates global ambiguity. The left column illustrates the “true” shapes of two facial profiles. The top right image is an image of the “squashed” face on the left, and the bottom right image is an image of the elongated face to its left. But these images in the right column are identical. How can this be?

Although there has been considerable progress in computer vision for certain problems such as object category recognition, no general-purpose algorithm can make decisions about objects in natural images with the flexibility and accuracy of a human--e.g. to determine not only identities, but object properties such as the material, the precise shape of parts and relations, or the precise locations of object boundaries.

From images to decisions and actions about objects: Preview of formalizing the problem

Formally, we want to understand how to get useful actions **A**, from image measurements **I**:

$$\mathbf{I} \rightarrow \mathbf{A}$$

Think of **I** and **A** as multivalued descriptions (e.g. vectors) of image measurements and action parameters. Actions are general--they could be motor actions such as reach and grasp, or descriptions of the fox above. To get from **I** to **A**, vision usually requires information about objects, surfaces, and scenes and their relationships to the viewer. Properties of objects and their relationships will be called **scene** attributes, represented by a vector **S**.

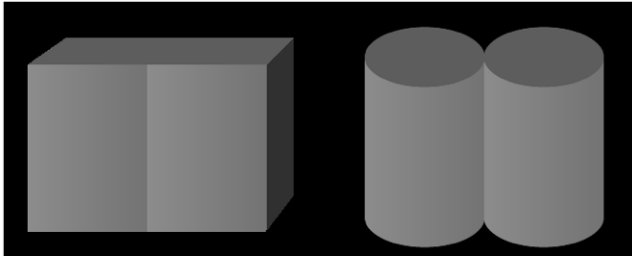
Consider the question: How can one estimate parameters of objects--their colors, shapes, materials, their relationship to other objects, to the viewer, to the viewer's hands, etc.--all from a glance? This is often referred to as the problem of *image understanding*, to emphasize that vision is a problem of perceptual inference. That is, given an image which is just a description of the light intensities at each point in space, and time (e.g. video camera or the sensors at the back of your eye), how can one infer the properties of the scene that caused the image?

In general, we'll represent the **image** by an array of intensities, $I(x,y,t)$, varying in space and/or time. (More generally, **I** could be some derived image measurements, like the location and orientation of local edge segments). Actions **A**, such as parameters required for grasping an object, or saying "that's a dog", require information about scene attributes **S**. Thus sometimes, we'll think of **A** as a function of **S**: $\mathbf{A}(\mathbf{S})$. But many times, we'll study problems in which $\mathbf{A} \sim \mathbf{S}$, i.e. in which an action parameter is the same as the scene property.

Then the problem is:

I -> S

An example is **S** = depth of an object from the viewer, or **S** = pigment property of the object. For example, the pattern of intensities on the left and right rectangles of the left block are the same. Further, the right pair of cylinders as the same pattern in the horizontal direction. Yet, the left rectangle appears darker than the right. But not so much for the cylinders. Why? The answer is that your visual system is trying to estimate the pigment level of the presumed surface being viewed--NOT the intensity. (example from: Knill & Kersten, 1991)



The problem of computing scene parameters from images is an example of an *inverse problem*. It is called this, because the goal is to estimate cause, such as change in pigment level, from data--the image measurements. Computing image intensities (the data) from the causes is called the forward problem, and involves specifying a "generative model".

The generative model

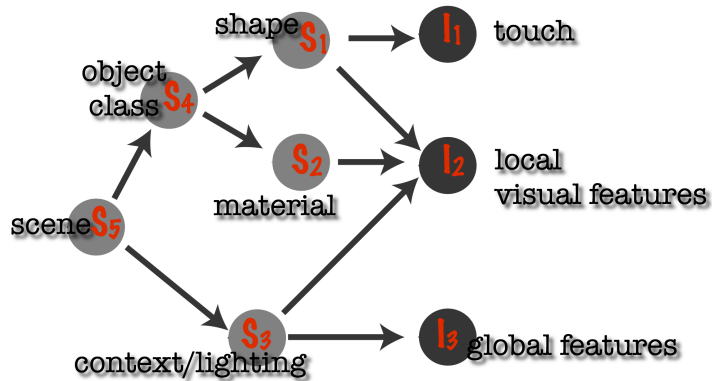
Visual understanding is only possible because images are a structured function of what is out there, i.e. images are functions of scenes: $I = \text{function}(S)$. Scene properties can be said to *cause* image patterns. This process can be captured by a forward or *generative* model.

3D computer graphics is a good example of a forward problem involving generative models. Sometimes called of "forward optics" -- how to go from a description of the scene environment to the image:

S -> I

In other words, a function that describes what "out there in the world" (object shape, lighting, etc.) causes the image intensities observed. This is also the problem of creating "virtual reality".

The causal structure leading to the data (image) is well-defined, providing a one-to-one generative model from scene description to image pattern. The figure below shows how one might specify a set of "knobs" **S** when synthesizing or generating stimuli **I** (which could include image, as well as other sensory patterns, such as tactile ones) using scene descriptions:



We'll spend some time learning both about methods for generative modeling of images, as well as techniques for solving vision problems by inverse inference.

Generative models are important because: 1) they represent the causes of the regularities that vision must exploit; 2) they can provide quantitative models for experimenters to manipulate visual stimuli to test models; 3) generative models may be used explicitly by the brain--a topic we return to at the end of the course.

Vision as (statistical) inference: reversing the generative model

So to sum up so far, the formal problem of vision is:

1. An image I has regularities caused by an underlying structure that has useful components S :

$$I = \text{function}(S)$$

2. The visual system has various goals (make decisions, estimates, descriptions, actions...) based on components in the underlying structure (S). I.e. to go from I to a description (or decision, estimate, etc.) S or more generally, an action A . We'll put a prime on S , i.e. S' to distinguish our estimate of a scene attribute from its actual value S :

$$I \rightarrow S'$$

This is the sense in which vision can be viewed as an inverse inference problem--inferring useful causes from image data. Methods of "good guessing" are needed because of the ambiguity problem--the inverse of the generative model can be one-to-many.

Distinguishing our estimate from the true value is important, because any estimator, including our own visual systems will make mistakes--we don't always see the correct depth. Further, we don't always see the same thing given the same image. Later we'll look at visual illusions that illustrate these points. We will see that the theory of statistical inference provides a natural framework to model under-constrained problems.

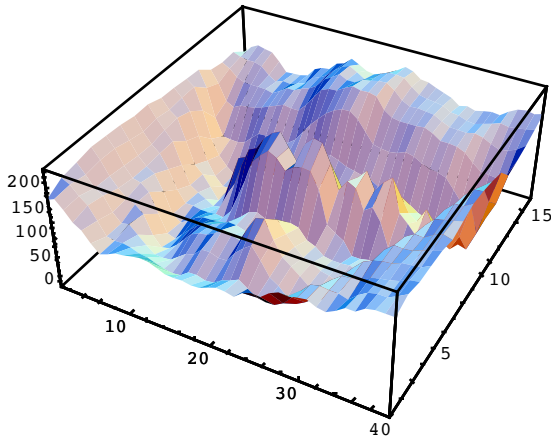
Over a century ago, Hermann Helmholtz described perception as "unconscious inference". Another way of expressing this is to view perception as a process that makes good guesses from indirect and ambiguous image input. As we proceed through the course, we will justify and amplify on the Helmholtz defini-

tion of perception.

More on the problem of ambiguity for objects

An example of ambiguity in depth and brightness: a "mystery image"

The following section has a "mystery" image I , which is just a 2D list (matrix) of light intensity values. It illustrates how several causes, in particular depth and material, get confounded in the image data. Let's make a plot that represents intensity as height:



This is a plot of I as a function of position, (x,y) .

Suppose our goal is to estimate depth at each position, i.e. obtain $S'(x,y)$ from I . One idea is to assume that I is proportional to depth, S at each location. Although naive, it isn't a terrible idea--there is some correlation between intensity and depth. Given this assumption, we'd conclude that the high middle ridge is closer than the bottom ridge. But as we will see in a moment, the dominant middle ridge that we see here does not correspond to near depths.

Consider another perceptual inference goal. What if we want to estimate the surface color (material, pigmentation or "paint") of the image? Let S'' be the surface color, where S'' is big for white surfaces, and low for black surfaces. Now it seems even more plausible that I would correlate very well with S'' . But how well?

Let's represent the mystery image in a form where your own visual system can judge...

Intensity representation of same data in mystery image:



First, note that the high intensity areas of the teeth are poor predictors of "nearness" in depth. They seem to be better measures of "whiteness". But what about the highlight on the lower lip? This produces a bump in the surface height plot above, but this a highlight due to the glossiness of the lip, and the pigmentation of the actual lip surface is not any lighter than other areas of the lip.

Information about shape and pigment is ambiguous, and the two are confounded in the simple image intensity measurements.

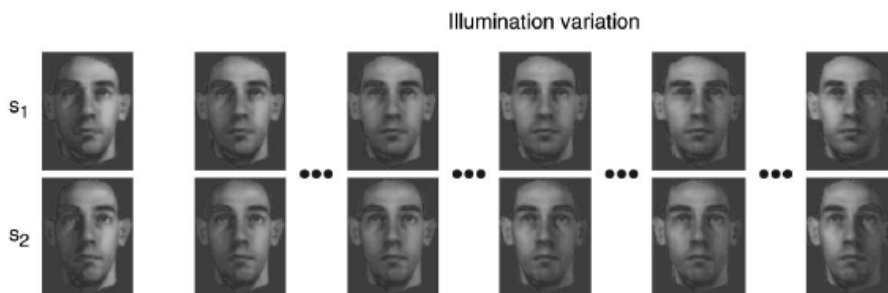
The problem of task

A recurring theme that will appear throughout this course is the importance of carefully characterizing the task of the person or "agent". Some kinds of scene or object information are more important to estimate the other kinds depending on the task of the agent, the behavioral goals.

Here's example of two tasks: Face identity vs. illumination direction. In the first task, given an image, decide whether it belongs to the George (s1) or Jim (s2).

Alternatively, there could be another task: given the same image, decide whether it is illuminated from the left or right.

The figure below illustrates the kinds of image variation that can result from different facial shapes (just two in this case), and different illumination conditions:



There are HUGE objective changes in the image going from left to right ("variability in illumination") that human perceptual judgments give very little weight to. The identity of a face is more important than which direction the light is coming from.

Where does the field of computational vision stand today?

It is only relatively recently that computer vision is obtaining practical solutions to the general object recognition problem: identify object categories in natural images from arbitrary viewpoints, illumination conditions, and in arbitrary contexts. Methods such as deep convolutional neural networks do remarkably well at learning given large annotated training sets. However, algorithms still lack the flexibility and robustness of human vision, and to generalize to instances that can be strikingly different from the training. Further, although there are similarities both human behavior and to the brain's neural networks, there are also substantial differences both in perception and in network details.

One of the major contributions of computer vision has been to define the mathematical problems of vision and to show that these can be quite difficult to solve. Historically, the problem of chess was considered a prototypical problem for Artificial Intelligence. Today we have machines that can beat most of us at chess. A lot of their power comes from high speed brute force search, likely quite different from the brain processes of the grand masters. Nevertheless, they can beat us, whereas the problem of picking out the chess pieces from the box (because of variability over viewpoint, lighting, material, and style) and setting them up on the board in the right places is still a formal challenge.

In this course we will study how the visual system deals with variability, such as due to over illumination, viewpoint and material. In addition to the problem that different illumination conditions (e.g. light source coming from above left or above right), and different viewpoints produce different images of the same object, vision has to cope with occlusion of one piece by another. Further, like fonts, different

chess sets have different styles. To some extent style variation can be modeled in terms of geometric variation. But the variations can be quite complex to characterize.

To further sober (and challenge you), the remarkable limitation of our understanding of visual inference is underscored by the fact that there are no machine systems that can solve the patently simple problem of deciding whether a surface has a light or dark pigment under general illumination. I.e., given a white or black chess piece, what color is it? The problem is that a black piece in bright light can have the same average intensity as a white piece in dim light. This is a problem that we will return to later in the context of human material and lightness perception.

On the positive side, there has been considerable theoretical and empirical progress in understand the problems of vision, how to solve them, and how the brain enables us to see with such remarkable competence. Hopefully this course will give you a useful and exciting introduction to the field. To get started, it helps to have an understanding of the interdisciplinary nature of vision science.

Computational vision: combining disciplines

Combining approaches from psychology, neuroscience, and computation

Vision is a part of cognitive science -- an interdisciplinary effort to understand the nature of knowledge, its acquisition, storage and utilization. It is also part of Cognitive Neuroscience—the study of the relationship between brain and cognitive behavior. I'd like to spend some time motivating the importance of an interdisciplinary study of vision.

Some motivation may be required here because of the nature of the course. In this course, we will study vision from a computational point of view. The topics are exciting because the course involves integrating knowledge across disciplines. But it can be frustrating because although it involves some math and computation, it isn't like most quantitative disciplines that have a structured sequence, and you will have to revisit content you've studied before, fill in gaps in your knowledge, and learn new concepts to make the interdisciplinary picture come into focus.

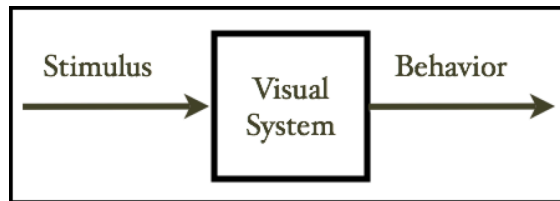
What is Computational Vision? It is the study of how to compute useful scene information, object properties, and action parameters from image measurements. And central to this course, it is in particular the study of how we as humans accomplish this computation. Thus, we will study human visual behavior to understand what information is and is not used for visual inference. And we will study neural systems because we would like to understand how the machinery of the eye and brain enables actions and the inference of properties of scenes in the world from images.

What does it mean "to understand visual perception"?

If we think we understand human vision, then we should be able to build or simulate a machine "which sees like we do". But what does this mean? There are several levels of abstraction that have to be considered, and the answer to this question requires careful thought. Robot vision, even if excellent, would not necessarily work the same way as human vision even at a very general level. For example, a vision system could rely on the reception of natural image patterns, or could actively send out signals to see how the environment modulates them. An example of the latter case would be to project alternating stripes of light and dark on surfaces, and then use the systematic distortion of the stripes in the image which to decode the shape. In fact, this latter principle is used (laser painted stripes) in commercial applications to measure shapes (e.g. Cyberware scans of the human face). But biological vision (in contrast to echo location) processes nature's patterns without active modulation of its images.

The study of cognitive science involves behavioral, biological and mathematical approaches. Let's take a closer look at the methods of three specific areas within each of these general approaches: Psychophysics, Neuroscience and Computer Vision. We will gain some familiarity with all three of these disciplines in this course, but let us first have a preview of their respective contributions and see how they relate to different levels of analysis.

Behavior and Psychophysics: Black-box approach to construct a model that "sees like we do"



This approach has a *major goal to qualitatively and quantitatively describe visual behavior*. Psychologists, ethologists and behavioral and systems neuroscientists all study behavior...but even physicists and mathematicians get in the act. Careful description is an essential first step in any science (e.g. Mendel, and genetics).

The study of behavior can be of at least two types:

First, we need to *know and understand the visual functions of an organism--its purpose and goals*. Human vision is used to identify objects, to read, to walk, to drive, to steady oneself, to reach and grasp, to throw, to plan, to judge beauty, and the list goes on. Different tasks require different kinds of image processing and inference.

The second type of behavioral study is *psychophysical analysis*. Psychophysics measures the behavioral consequence of physical (or informational) variations in the image stimulus with a goal towards understanding underlying neural mechanisms. Examples are: the measurement of apparent brightness as a function of physical light intensity; just discriminable differences in light intensity; changes in sensitivity as a function of adaptation; changes in recognition performance as a function of viewpoint. Clever psychophysical experiments can reveal not only the diverse visual processing requirements, but also test hypotheses about alternative accounts of a given process.

Psychophysics goes beyond mere description and historically has made some striking predictions about the nature of the underlying biology. For example, the psychophysics of color matching in the 19th century anticipated three physiological cone receptor types and their relative spectral sensitivities as a function of wavelength. This so-called "trichromacy" theory of was not physiologically established at the cell level until the 1960's. We will see later how psychophysics in the 1940's showed that photoreceptors (rods) in the eye could transduce single photons into an electrical signal. Certain brightness illusions discovered in the 19th century suggested patterns of neural connectivity and spatial interaction (called "lateral inhibition") that were not put on a firm physiological and neural foundation until the 1950's.

A psychophysical approach has clear limits in its ability to give an account of how we see. One reaches a point where too many theories of what is "inside the box" give the same input/output relation in the psychophysical data. For example, computing y as: $y = x(x - 1)$ gives the same mathematical relationship as a different computation in which x is subtracted from its square: $y = x^2 - x$. The fact

that different combinations of wavelengths of light appear the same could have many neural explanations. Researchers eventually "go inside the box", an animal "model" of the box (like a frog, cat or monkey) to find out what was going on at a finer level of analysis. This brings us to the methods of neuroscience, such as anatomical tracing, electrophysiological recording from single neurons, and in humans, brain imaging.

Neuroscience: Going inside the box

What happened when physiologists and anatomists looked at the biological basis of the psychophysical descriptions? Indeed, as discovered using microspectrophotometry of single cones in the 1960's, there are 3 distinct types of cone photoreceptors in the retina at the back of the eye. Electrophysiological recordings showed that their spectral sensitivities were remarkably similar, but not identical, to those inferred from psychophysics. It was discovered that are neural circuits (lateral inhibition) in the retinas of horseshoe crabs, cats, and monkeys that behave like Ernst Mach predicted to account for certain brightness illusions.

And as we will see in the next lectures, 1940s psychophysical predictions were by verified neuroscientists in the 1970's: when rod photoreceptors were sucked up into tiny glass tubes and exposed to extremely small amounts of light, they could be seen to transduce single photons. Later, we will see examples of more recent neural accounts of psychophysical observations that go beyond retinal processing to other parts of the brain, accounts that are being tested using both electrophysiological and brain imaging techniques.

In the 1950's and 1960's there was a tremendous excitement that we could understand the brain's function and in particular visual perception in terms of single neurons....but neurobiologists had probably been particularly lucky...at least they were more fortunate than if the brain had been designed like a modern digital computer. In the 1970's, it became increasingly apparent that understanding how biological systems discriminated light and pattern was only scratching the surface of the problem of vision. Computer vision was beginning to show that competent vision was truly a problem of sophisticated inference and estimation. The number of visual areas discovered in the cortex of the brain grew.

Vision was becoming more complicated and harder than expected.

Computer Vision: The need for tools to deal with the complexities of perceptual inference

Imagine the following example. Sometime in the distant future, Martian scientists have acquired a Terran computer device that plays an ancient video game, say Super Mario Brothers. Now consider the various ways these scientists might go about trying to understand this device.

First, they could adopt technique adapted from a neuroscience, "anesthetize" the computer (i.e. just take away the screen, so there is no external output), and begin using a volt meter or a logic probe to figure out what the box is doing. This is like doing neuroscience without psychology or psychophysics. The scientists might learn about logic gates, shift registers, and RAM, etc.. But, what are the chances of figuring out that the machine was even designed to play a game? Pretty slim.

But now give the scientists a working system complete with screen and the controls, but minus the logic probe. This is like doing behavioral science--psychophysics. With some careful experiments, they could begin to figure out the rules of the game. (Although, they might be left with questions forever unanswerable, like "why was this machine built in the first place?"). If asked to "build a machine" that does the same thing, the Martian scientists might still have a hard time. They might be able to build a

copy that mimics the behavior of the game, but even if the scientists had a solid body of results using the logic probe and observing the functioning system as a whole with the screen on, there is still something missing. They would have missed the point that the essential structure of the game is not the hardware, nor the input-output relations, but rather a highly complex computer program. They need an understanding of the the algorithms, and how that these are related to the hardware that supports it.

In short, what is missing is an understanding of how the pieces fit together to solve a specific information processing task-- a task that involves getting magic mushrooms, escaping turtles, smashing brick ceilings to get coins, jumping up flag poles, etc. Without this knowledge, they would be unable build new video games, like Mario Brothers. True understanding of vision should result in the generalization, e.g. the capability of building a machine that sees like us, but which may differ from the original in ways that we can understand.

This example illustrates the need for a **computational approach to vision**. The computational level involves understanding how image patterns are formed (generative models), and how scene inferences can be drawn from image patterns. To handle the complexity of natural patterns requires the tools of computer programming, and has been increasingly emphasized in recent years, the mathematics of statistical inference.

Although the computational approach grew out of the early communications theory, cybernetic, and artificial intelligence studies of the late 1940's, 1950's and 1960's (e.g. Turing, von Neumann, Wiener, Shannon), one of the chief protagonists of this approach for the study of visual perception was the late David Marr from MIT in the early 1980's.

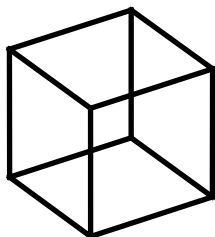
Computational vision requires study at several levels of analysis

Although the MIT scientist, David Marr made a number of specific contributions to understanding human vision in the late 70s early 80s, he is particularly well-known for his elucidation of a computational approach to vision as necessarily involving of multiple levels of explanation (Marr, 1982). Let's see what this means.

Functional ("computational") Theory. What is the goal of a computation?

Why is it appropriate? What strategy can carry it out? Both psychology and theoretical analysis help to answer these questions. For example, the Necker cube (below) perceptually flips because the goal of the visual computation is to represent the 3D structure of the objects causing the 2D retinal image. There are two equally plausible 3D interpretations. But are there only two?

Necker cube

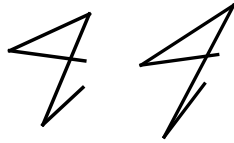


Representation and algorithm. How to represent input and output? How to get from input to output?

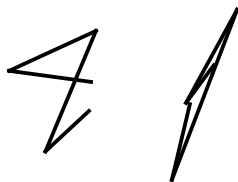
Psychology and computer program models help answer these questions. For example, mental rotation experiments done with human observers give clues as to how visual information is represented and

processed. The time to decide whether a sample figure on the right is a rotated version of the one on the left or not, increases monotonically with the actual angle required to check the match (over a certain range and conditions).

The right figure is an image of the same object as the one on the left, but is rotated by 30 degrees about the vertical axis:



In the next figure below, the right figure is rotated by 80 degrees:



Implementation or hardware. How to build it with actual components?

An algorithm can be implemented with pencil and paper, devices with mechanics, vacuum tubes, silicon chips, or neurons. Neurobiology and neural network computer simulation help to understand the particular characteristics of neural computation. For example, after-images are a well-known perceptual phenomenon. The after-image of a flash from a camera has to do with how human vision implements transduction in the retina, and the receptors in particular.

In future lectures we will see the relationships between human behavior, neurophysiological mechanisms, and computational theory. Sometimes the computational theory comes first, but sometimes we will work backward from an experiment or visual phenomenon to the theory. For example, we will look at:

- o quantum limits to vision-- What are the theoretical limits to light discrimination? -> Computational theory of discrimination.

- o lateral inhibition in neural networks-- for detecting edges? or to reduce redundancy? -> Computational theory of neural image coding.

The next lecture will begin by studying one of the simplest of vision problems: How well can we detect and discriminate light intensity? What are the limits to this ability? Quantum nature of light? What is the computational theory for brightness discrimination?

Getting started with Mathematica

Mathematica vs. Matlab vs. Python (Jupyter notebooks)

Front-end and Notebooks: Organize, outline, document. Program, evaluations, data all in one place

Kernel: Separate program does the calculations

Go to the Help menu in *Mathematica*. Go to Documentation Center, and from there to Getting Started Videos or Find Your Learning Path.

To get an overview with audio, go to the screencast:

<http://www.wolfram.com/broadcast/screencasts/handsonstart/>

<http://www.wolfram.com/support/learn/get-started-with-mathematica/>

You can read *Mathematica* files free with Mathematica CDF Player.

Introduction

See Assignment #1 in the course web page.

References

- Helmholtz, H. v. (1867). Handbuch der physiologischen optik . Leipzig: L. Voss.
- Hoffman, D. D. (1998). Visual Intelligence . New York: W. W. Norton & Company.
- Kersten, D. High-level vision as statistical inference. (1999) The New Cognitive Neurosciences, 2nd Edition, Gazzaniga (Ed.). MIT Press.
- Kersten, D., & Yuille, A. (2003). Bayesian models of object perception. Current Opinion in Neurobiology, 13(2), 1-9.
- Knill, D. C., & Richards, W. (1996). Perception as Bayesian Inference . Cambridge: Cambridge University Press.
- Marr, D. (1982). Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. . San Francisco, CA: W.H. Freeman and Company.
- Mumford, D. (1995). Pattern theory: A unifying perspective. In D. C. Knill, & R. W. (Ed.), Perception as Bayesian Inference (Chapter 2). Cambridge: Cambridge University Press.
- Poggio, T. (1984). Vision by Man and Machine. Scientific American, 250, 106-115.
- Zeki, S. (1993). A Vision of the Brain . Oxford: Blackwell Scientific Publications.