Goals of the Course

- To give an introduction to the state of the art computational models of the mammalian visual cortex.
- To describe the relevant evidence from anatomy, electrophysiology, imaging (fMRI), and psychophysics.
Course Materials

- Latex Notes (D. Kersten and A. Yuille).
- Papers – reading list.
- Handouts.
- Readings: “Understanding Vision: Theories, Models, and Data”. Li Zhaoping.

- Grading: Homework Assignments and Course projects.
Structure of this talk

(I) Introduction
(II) The Visual System and the Brain
(III) The Complexity of Vision
(IV) Vision as Inverse Inference
(V) The Visual Hierarchy
Part I: Main Points of this Talk

(I) Vision is extremely hard. Humans are vision experts.

(II) The Visual system is very complex and only partially understood.

(III) Why is vision hard? The key problems: complexity, and ambiguity.

(IV) Visual hierarchies.
The Purpose of Vision


Information Processing: receive a signal by light rays and decode its information to understand the scene.

Vision appears deceptively simple, but this is highly misleading.
Humans can understand very complex images.

- We can get the “gist” of an image in about 150 msec, the time to blink an eye.
But we do make mistakes

- Perception is not reality.
We are fooled by accidental alignments
Part II: Vision and the Brain

- Humans appear to understand images effortlessly. But this is only because of the enormous amount of our brains that we devote to visual tasks.

- *It is estimated that 40-50% of neurons in the cortex are involved in doing vision.*

- Humans are very visual. We get much more information from our eyes than other animals.
The retina and optic nerve

- Images are captured at the retina.
- They are transmitted to the visual cortex by the optic nerve.
The Retina, LGN, and Visual Cortex

(I) Visual Information received at Retina.
(II) Information is sent along the Optic Nerve.
(III) It passes through the LGN.
(IV) Information is decoded in the Visual Cortex.
The Visual Cortex

- Vision/perception is performed in the visual cortex. This is organized hierarchically.
- The visual cortex is roughly 40% of the entire cortex.
Nerve Cells and the Brain

- The brain is made up of two types of cells.
- Neurons and Glial Cells.
- Neurons are the basic elements of the nervous system. Information travels along them by electric impulses.
- Neurons transmit to other neurons by synapses.
- Glial cells play a supporting role. Maintain brain structure. Do not conduct electricity.
Nerve Cells and the Brain

- Neurons and Networks of Neurons.
- Brain has 100,000,000,000 neurons.
- Each communicates with several 1,000 other neurons.
Nerve Cells and the Brain

- The number of nerve cells 100,000,000,000 is about the number of trees in the Amazon rainforest.
- The number of nerve cell connections 1,000,000,000,000,000 is about the number of leaves in rainforest.
- The number of connections in the world’s telephone system (biggest machine on the planet).
- The brain is the most complicated (known) system in the Universe.
Caveat

- We only know the neural wiring diagram for the simplest creatures – e.g., C Elegans.
- There are large initiatives to study the wiring diagrams of more complex animals (e.g., mice).
- Neurons themselves are complicated. More complicated (much more?) than “artificial neural networks”.
How to study the brain

- Exciting new methods, e.g., Connectome, Optogenetics.
Neural Prosthetics?

- Patient is paralyzed but can still think, or plan, the actions.
- Implant electrodes in the “planning area”
- Decode the neural activity and thereby enable a patient to control a robot limb.
- Proof of concept. Monkeys moving computer cursors.
Functional Magnetic Resonance Imaging (fMRI)

- Use fMRI to measure blood flow in the brain.
- Active areas require more blood. fMRI can measure the oxygen content of the blood.
- Molecules in the blood cells respond differently to the magnetic field depending on how much oxygen they are carrying.
Part III: Why is Vision hard?

- The main difficulty of vision is due to complexity.
- Complexity and ambiguity of stimuli. The complexity of the visual tasks.
- First realized when Artificial Intelligence researchers tried to make computer vision systems.
A Brief History of Artificial Intelligence and Vision

- Initially AI researchers thought vision was easy. “Solve vision in a summer”. (1966).

- But the difficulty of vision rapidly became clear as researchers tried to get computer programs to interpret images. Nothing worked.

- Researchers started realizing that vision was much harder than “intelligence tasks” like playing Chess.
Why is Vision Hard?

- Look at the raw input displayed as a set of numbers which plot the intensity as a function of position (bottom left).
- The images are very complex. They are of the same bike and tree. But they look very different.
Complexity of Images and Visual Scenes

- The set of all images is practically infinite.

- Only a tiny fraction of all possible images have been seen by humans.

- The number of visual scenes is also enormous. There are 30,000 different types of objects. They can be arranged into 1,000 scene categories. This can be done in an exponential number of ways.
Humans can estimate rich descriptions from images.

- Humans can easily detect the fox, the tree trunk, the grass and the background twigs.
- And can also estimate the shape of the fox’s legs and head, its type fur, what it is doing, is it old or young, is this winter or summer?
- But the local regions are highly ambiguous.
But local regions are ambiguous
Local ambiguity

Figure 4. The outlined region around the boundary of the log (left panel) is shown expanded in the middle panel. The right panel shows how a hypothetical array of model V1 neurons (Gabor filters at four different orientations) would respond to the image subregion shown at left. The length of each line segment indicates the magnitude of response of a neuron whose receptive is situated at that position and orientation. An array of such neurons provides only weak or ambiguous cues about the presence of object boundaries in natural scenes.
The Fundamental Problem of Vision?

- How can a visual system deal with this complexity?
- How can humans represent, learn, and rapidly access (infer) the structure of the world?
- How can we do it so quickly?
- “The next century will be the century of complexity” S.W. Hawking.
Part IV: Vision as Inference

- Helmholtz. 1821-1894.
- “Perception as Unconscious Inference”.

![Image of a stamp with a portrait of Hermann von Helmholtz]
Vision as Inverse Inference

- Images are generated by highly complex processes – light bounces off objects and is captured by the retina/camera.
- This process is studied in Computer Graphics. It models objects, light sources, and how they interact.
- *Vision must invert this image formation process to estimate the “causal factors” – objects, lighting, and so on.*
Inverse Problems are hard

- There are an infinite number of ways that images can be formed.
- Why do we see a cube?
- Prior – cubes are more likely than other shapes consistent with the image.
Bayesian Decision Theory

- Bayes’ Theorem gives a conceptual framework to solve inverse inference problems.
- It states that we can infer the state $S$ of the world from the observed image $I$ by using prior knowledge.
- $P(S|I) = P(I|S)P(S)/P(I)$.
- Rev. T. Bayes. 1702-1761
Neural networks and Machine Learning

- Bayes gives a conceptual framework for visual perception. But it is far from being sufficient.

- *Neural network and machine learning methods have become increasingly effective.*
Inverse inference requires priors

- Humans use prior knowledge about the world (obtained through experience).
- This usually gives accurate perception. But sometimes fails (e.g., change blindness).
- Visual illusion give insights into visual strategies and prior assumptions.
The Ames room

Which girl is bigger? Trick question.
Perspective Illusions
Lighting Illusions: which square is brighter?
Visual Illusions and Context

- The perception of brightness of a surface, or the length of a line depends on context.
- Not on basic measurements like:
  - the no. of photons that reach the eye
  - or the length of line in the image.
- Humans perceive images by making assumptions about the structure of the scene. Illusions arise when these are wrong.
Flying Carpets?

- Can people fly? You think that the shadow is cast by the rug the woman is standing on.
- But instead it is cast by a flag outside the picture.
Levitation?

- Accidental alignment.
- The wet spot on the ground is misinterpreted as a shadow.
- This shadow is usually at the contact point of the human and ground.
Part V: Visual Hierarchies

- The human visual system is arranged in a hierarchy.
- This is a natural way to deal with the complexity problems of vision. Can it be derived from first principles?
- It also gives a way to roughly taxonomize visual problems.
Feedforward Models

- Multiple stages of processing which build progressively more complex/abstract representations.
- This mimics knowledge of the Ventral Stream – V1, V2, V3, V4, IT.
- Hierarchical organization – where V1, … V4 all have a map of viewed space (centre).
Hmax: Poggio et al.

- Invariant representations of objects are built up though a hierarchical feedforward algorithm.
- Each stage is composed of separate populations of neurons that perform feature extraction and spatial pooling.
Bio-Inspired Models

- Deep Belief Networks (Krihevsky et al. 2013)
- Image classification – ImageNet.
- Sharing of hierarchical features.
Visual Tasks

- But current hierarchical models do not address the complexity of visual tasks that humans can perform.
- We do not just classify objects, we detect object parts, estimate three-dimensional structure, reasoning about it, etc.
- Vision is part of an action perception loop.
Figure 2: Components of scene analysis. The scene itself contains not just a single target object, but other objects, terrain, and background, all of which may be important for behavior. The neural structures enabling scene analysis contain multiple levels of representation and analysis. The level of “intermediate features” is where inferential processes come into play. (From Lewicki, Olshausen, Surlykke & Moss, 2013)

- Marr proposed that vision is broken down into constructing a sequence of representations.
  - (I) A primal sketch – represents features and tokens from the image.
  - (II) A 2.5-D sketch that makes explicit aspects of depth and surface structure.
  - (III) An object-centered 3D model representation of objects.
Low-, Mid-, and High-Level Vision

- Vision can be broken down into low-, mid-, and high-level vision (very roughly).
- Low-level vision – local image operations which have limited knowledge of the world.
- Mid-level vision – non-local operations which know about surfaces and geometry.
- High-level vision – operations which know about objects and scene structures.
Low-level vision

- Image processing.
- Filtering, denoising, enhancement.
- Edge detection.
- Image segmentation.
- Right: ideal segmentation (followed by labeling).
Mid-Level Vision: Depth

- Estimation of 3D surfaces:
  - E.g., binocular stereo,
  - structure from motion,
- Figure: Images, Depth,
- Segmentation.
- (Blue close, red far).
Mid-Level Vision: Grouping

- Kanisza. Humans have a strong tendency to group image structures as surfaces which can partially occlude each other.
High-Level Vision

- Object detection and Scene Understanding.
- Example: detect objects and object parts.
Mumford (1991) argued that the feedback and feedforward connections suggests that the visual system uses generative models.


Figure 7. Hierarchical Bayesian inference. The variables represented at each level are inferred from a combination of bottom-up and top-down inputs. Bottom-up inputs enter into the likelihood, while top-down inputs enter into the prior. The two are combined to form the un-normalized posterior, which guides the inference of variables at each level.
Feedforward and Feedback

- Bottom-up proposals are validated by top-down generative models.

Hierarchical Compositional Models

- Compositional models represent objects and scenes in terms of compositions of elementary shared parts.
- This offers a possible solution to the complexity problem of vision.
High-Level Vision and AI

- Understanding of objects, scenes, and events.
- Reasoning about functions and roles of objects, goals and intentions of agents, predicting the outcomes of events.
- The full Artificial Intelligence problem.

Figure 1. Example of image understanding. Analysis of the image (top-left) produces a parse graph (right) representing hierarchically objects, contextual relations, and semantic associations (in italic orange font) for attributes, functions, roles, and intents. The parse graph maybe converted to a description in natural language (bottom-left).
Vision is an extremely difficult problem. It seems easy to us because our brains are specialized to perform visual tasks.

The key challenge of vision is complexity. Complexity (and ambiguity) of images, the world, and the visual tasks.

The visual system is a large part of the human brain, which is the most complex physical system we know.
Vision is inverse inference. Estimating the three-dimensional scene from input images.
The visual system is organized into a hierarchical structure. This includes feedforward and feedback processing.
This hierarchy leads to a rough taxonomy into low-, mid-, and high-level vision.
Some readings

- B. Olshausen. Perception as an Inference Problem.