Yuille & Kersten: Early Vision

Rachael Steiner

Feb. 2, 2017

What is vision?

Taking light that hits the retina and turning it into information.

Why is it interesting?

- Images are highly complex, dynamic
- We can rapidly identify objects, their materials, movement, and orientation.
- Visual system detects and makes use of **statistical regularities** in the input that correspond to regularities in the world.

Remember Marr?

- Vision is studied at **behavioral, neural, and computational levels.**
- Computational models focus on either *understanding how* humans and other animals see, or *replicating the outcome* without worrying about the process.

Coping with complexity: Simplifying stimuli and neural circuits

- A 100px by 100px image with 256 possible color values has 256^10,000 possible manifestations.
  - That's a lot.
- Instead, normally use synthetic stimuli with only task-relevant information
- Also helps to isolate the effect
- But still need to think about generalizability: humans usually perceive those complex images.
- Similarly, need to use models with fewer neurons, less complex firing patterns.

Breaking it into tasks

- Visual processes often modeled as individual tasks (known as **modules**).
- Output **representations** which can be fed to other modules.
- Requires caution: probably not all tasks are independent.

Grouping tasks into levels

Low-level tasks:
Estimating local properties, finding object boundaries, estimating motion flow

Mid-level tasks:
Estimating properties of surfaces (shape, texture, position), depth ordering

High-level tasks:
Estimating properties of objects, relationships among objects, actions, structure of the overall scene.

Low- and mid-level vision are together referred to as **early vision.**

Low-level vision

Processing that can be done without explicit world knowledge

- Mostly for finding differences: edge detection, segmentation
- Estimates motion by comparing intensity (brightness) across images
  - Relies on statistical regularities
- Not actually that good at edge detection on its own: probably suggests a *possible set* of edges which can be narrowed down by higher object models
  - More reason for caution when studying levels in isolation.

Mid-level vision

Processing that "knows" about geometry, materials, lighting, but not objects or scene structures

- **Perspective projection:** uses **vanishing points** and the assumption that there is likely a ground plane to determine the orientation of a surface
- Knows that objects can partially occlude each other —> **depth ordering**
  - Further supported by binocular vision: can technically use trigonometry to estimate distance after establishing a correspondence between two images
- Depth and shape from shading, textures, contours
- Object properties from textures: A shiny patch of ground might be icy —> Don't step on it.

In the brain

- Retina —> lateral geniculate nucleus —> visual cortex
- Light activates **photo-receptors** in the retina
- **Ganglion cells** (also in the retina) handle output from the photo-receptors and encode it for transmission via the **optic nerve**

So the retina has two main challenges:

1. How to cope with images of highly variable intensity
2. How to encode images for fast but robust transmission

Dealing with variability

- Range of intensity is far greater than neurons can encode
- Theorized that the ganglion cells actually only encode local contrast

I wonder if the pupil helps at all to reduce the range too?

Information transfer

- The optic nerve isn't that big, so how is data reduced for transfer while somehow maintaining its fidelity?
- Some help from information theory, statistical knowledge of the stimuli

Could it be that the ganglion cells mostly only encode relevant changes? (like a `git diff`)

Visual cortex

Often divided into two streams:

- **Ventral stream:** "what" — object detection, scene understanding
  - V1, V2, V4, inferior temporal regions of the extrastriate cortex.
- **Dorsal stream:** "where" — analyzing movement and position
  - V1, medial temporal cortex, parietal cortex
- Probably actually more complex than this

Summary

- Vision is a highly complex process
  - Patterns of light intensity are converted into information
  - Simplifications are helpful, but need to be taken with caution
- Vision happens more or less hierarchically, with each process building on the last
  - But there is likely some amount of interaction, rather than being purely feed-forward.