Lecture 4
High-level vision:
Recognition, learning

International Summer School and Workshop on Vision,
Learning and Cognition

August, 2012

Daniel Kersten
kersten.org
Main points so far...

Perception has “built-in” knowledge of generative structure of images
Perceptual “explaining away”

This is a bi-stable percept
The stimulus is always the same. But when your brain organizes the four lines into a diamond shape, activity in V1 goes down.

Multi-stable perception

Lorenceau & Shiffrar

Fang, Boyaci, Kersten, Murray (2008)
Results consistent with feedback down the visual hierarchy

We considered two possibilities:

- neural representations of hypotheses ("diamond shape") in higher visual areas suppress activity representing consistent features at lower level
  - "shut-up"

- neural representations of hypotheses ("diamond shape") in higher visual areas enhance activity representing consistent features at lower level, while suppressing inconsistent features
  - "stop gossiping"
• We also noted behavioral results showing that human vision combines multiple sources of information (local measurements & priors, multiple local measurements) weighted by reliability

• Suggests
  – the neural code may represent probabilities (e.g. not just central moments, but also higher moments) and neural computation takes uncertainty into account

• Bayesian brain
• What features do humans learn to use when categorizing objects?
  – discriminative learning

• How do humans learn object models?
  – generative model learning

• Patterns of brain activity during recognition in clutter
Learning informative features for a task

What do these scenes have in common?

With Evgeniy Bart & Jay Hegdé
“Up” curbs-- that require a step up
Distinguish from non-“up curbs”

...that do not require a step

But may require a different action
Learning based on informative fragments for the task

- Find fragments that maximize mutual information for class membership (Ullman et al., 2002; Bart et al, 2004)

- Detect “up curbs” from an approach angle that requires a step
Learning object categories

Do image features (fragments) that maximize mutual information predict the features that human observers learn to use?

Need novel object classes with small within-class variation and slightly larger between-class variation

Virtual phylogenesis of digital embryos


Virtual Phylogenesis

Icosahedron

Step 0

G₀

Step 1

G₁

Step n

Gₙ

Shape Class A

Shape Class B

Shape Class C
Training

A

A or B?

B
Testing

Sample Object  Test Object  Sample Object

A  Main Fragments  C  Control Fragments

0  1  2  3  4
5  6  7  8  9
Results
Lecture 4

• What features do humans learn to use when categorizing objects?
  – discriminative learning

• How do humans learn object models?
  – generative model learning

• Patterns of brain activity during recognition in clutter
What limits perceptual ability to parse an image, e.g. to separate figure from ground?

- Objective information
- Local diagnostic features
  - edges, color, texture, motion, stereo
- Global object knowledge
  - prior knowledge, expertise with shape classes (e.g. faces, birds, ...)
- Neural mechanisms of the observer
Knowledge required to resolve ambiguity

Impossible | Hard | Easy

Segmentation difficulty

High-level knowledge | Local features

Importance of prior knowledge

Camouflage

Computational solutions?

- **Bottom-up**
  - **Finding object contours**
    - *Simple edge detection algorithms don’t distinguish important edges from the unimportant ones*
  - **Finding regions**
    - *Grouping based on similar textures helps, but still has problems with occlusion, etc.*
  - **Better features—motion, color, stereo help, but...**

- **Top-down?**
  - **Analysis by synthesis?**


...but there is a problem with analysis by synthesis

- How to recognize/segment an object in clutter (or camouflaged) without a prior model of its shape?
- Can the brain build a model of the object without seeing uncamouflaged (e.g. low-ambiguity or segmented) examples?

*This is the “bootstrap learning problem”*
Solutions?

• Learning to identify and segment is “opportunistic”

• Learning a novel object takes place only when image information is sufficiently reliable to segment object from background…when difficulty level is “easy”

• e.g. target color or motion enables easy segmentation
An experiment:
Learning to recognize novel objects in clutter

- Capture some aspects of natural object learning conditions
- Novel object synthesis
- Image variation model
  - Camouflaged, unoccluded, yet “invisible” target objects
- Clutter composed of similar objects
- Training presentation for learning

Want novel objects, yet have some of the shape regularities of biological form

Again, make novel objects using virtual morphogenesis -- synthesize “Digital embryos”
The “paint job”

Old style

New style

Introduces lots of local false positives by mimicking shading at occlusions, curvature, folds
Callionima
Make target object and background have similar texture statistics
Training sample:
Target object in “plain view”
How good is the camouflage?

Typical attempt failed to find and trace outline.
Training: 3 “clue” conditions

- Two easy conditions:
  - Color
  - Motion -- “very easy”
- One hard -- completely camouflaged
  - “No clue”, static
- Camouflage designed so that:
  - objects in any two scenes could not be reliably put in register by correlating
    - image intensities, or
    - extracted edges
  - No individual training image could be segmented by a human observer
No clue: sample training session
Did you learn any objects?
Color clue
QuickTime™ and a MPEG-4 Video decompressor are needed to see this picture.
Training details

3 objects, (A,B,C), to learn
Each object is presented, in “camouflage” 5 times each day for 4 days
– 10 sec presentation
– Unique sound associated with each object

Fixed camera and lighting, but *new camo’ pattern, random translated position, and new background for each presentation*

3 “clue” conditions (one each week over three weeks)
Testing

• Identification

• Test 24 hours after each training session

• Test object was a freshly camouflaged training object and background: (A,B,C) or new

• Choose A, B, C or new

• Tracing on last day

• Trace shape contour outline of training object with new camo and background
Some people failed to learn to identify the camouflaged objects, when trained with camouflaged objects.
But others did learn to identify the camouflaged objects
Green: Observer’s correct trace   Red: Observer’s extra traces
Yellow: Missed traces
Perfect trace after training
Global position error
Apparent global scale error
Missing parts
Conclusions

• Humans can quickly learn object models from training examples in cluttered images that individually cannot be parsed without prior experience

• They apply these models to do accurate segmentation

• We call this bootstrap learning

• But people make mistakes
  • consistent with learning parts
Brain representations?

Neuroimaging results

Measure patterns of brain activity while human observers are recognizing digital embryos in clutter under two conditions:

- prior opportunistic learning
  - i.e. easy to segment during training
- prior bootstrap learning
  - hard to segment during training
Same test in clutter--different segmentation cues during learning

Conclusion

Very different patterns of brain activity during recognition depending on how the objects were learned

Different features?
Different mechanisms?
Summary

- Human observers can discover diagnostic features object features for categorization
  - consistent with learning features of “intermediate complexity”
    - V4?
- Human brain can build “object models” given highly ambiguous training images-- *bootstrapped learning*
  - exactly what is learned and when?
  - *upcoming talks, discussion on learning in hierarchies*
- Pattern of brain activity during recognition in clutter depends on how objects were learned