Visual Object Recognition
Computational Models and Neurophysiological Mechanisms
Neurobiology 130/230. Harvard College/GSAS 78454

Web site:  http://tinyurl.com/visionclass  (Class notes, readings, etc)
Location:  Biolabs 1075
Time:  Mondays 03:30 – 05:30
Dates:  Friday 09/04*, Mondays 09/14, 09/21, 09/28, 10/05, 10/19, 10/26, 11/02, 11/09, 11/16, 11/23, 11/30, 12/07*

Lectures:
Faculty:  Gabriel Kreiman and invited guests

Contact information:
Gabriel Kreiman
gabriel.kreiman@tch.harvard.edu
617-919-2530
Office Hours: After Class. Mon 05:30-06:30
Class 1. Sep-04  Introduction to pattern recognition. Why is vision difficult?


Class 3. Sep-21  Psychophysics of visual object recognition [Ken Nakayama]


Class 5. Oct-05  Introduction to the thalamus and primary visual cortex [Camille Gomez-Laberge]

Oct-12  Columbus Day. No class.


Class 7. Oct-26  First steps into inferior temporal cortex [Carlos Ponce]

Class 8. Nov-02  From the highest echelons of visual processing to cognition [Leyla Isik]


Starting from the very beginning

• Let there be light, and there was light.

• Objects reflect light

• Light photons impinge on the retina

• The retina conveys visual information to the brain

An oversimplified first-order description:

*The retina functions as a very sophisticated and spectacular digital camera*
Natural images are special
We only encounter a small subset of the space of possible images

• Consider an image with 256 shades of gray per pixel
• Consider an image of size 100 x 100 pixels
• How many such images are possible?

Answer
For a size of 1x1 pixel, there are 256 possible images.
For a size of 1x2 pixels, there are $256^2$ possible images.
For a size of 100x100 pixels, there are $256^{10000}$ possibilities*

Yet, we only encounter a small fraction of these possibilities in natural images

*Some of those are “related” by translation, rotation or inversion, etc
Spatial regularities in natural scenes

The properties of nearby points are correlated

Figure 3  (a) Joint distributions of image pixel intensities separated by three different distances. (b) Autocorrelation function.

Simoncelli and Olshausen 2001
Images contain different spatial frequencies
Some are more prevalent than others

Low frequencies  Middle frequencies  High frequencies

Natural image statistics

Power spectrum $\sim 1/f^2$

\[ \log(f(w)) = \alpha \log(w) + c \]

Note: Scale invariance

\[ w' \rightarrow aw \]

\[ \log(f(w')) = \beta \log(w) + d \]

There are multiple examples of power law distributions in physics, biology and social sciences

Figure 4  Power spectrum of a natural image (solid line) averaged over all orientations, compared with $1/f^2$ (dashed line).

Simoncelli and Olshausen 2001
Natural image statistics
There are also strong correlations in time

The visual input is largely static, except for:
  • External object movements
  • Head movements
  • Eye movements

The visual image is largely static over hundreds of milliseconds

Silent Reading    225-250 ms fixation, 2 degrees saccade size (8-9 letters)
Scene Perception  260-330 ms fixation, 4 degrees saccade

“Slowness” has been proposed as a constraint for learning about objects (Földiak 1991, Stringer et al 2006, Wiskott et al 2002, Li et al 2008)
Example pattern of fixations during a movie
Microsaccades are important for perception

The eye constantly makes very small “fixational” movements, a.k.a. microsaccades.

When visual stimulation is held fixed relative to the retina, visual perception rapidly fades (Ratliff & Riggs, 1950).

These microsaccades may be critical for counteracting perceptual fading (McCamy, 2012)

Martinez-Conde, Macknik, & Hubel, 2004
An image as a collection of pixels

<table>
<thead>
<tr>
<th>57</th>
<th>53</th>
<th>58</th>
<th>63</th>
<th>44</th>
<th>41</th>
<th>66</th>
<th>93</th>
<th>68</th>
<th>25</th>
<th>67</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>52</td>
<td>117</td>
<td>130</td>
<td>121</td>
<td>124</td>
<td>119</td>
<td>130</td>
<td>94</td>
<td>34</td>
<td>58</td>
</tr>
<tr>
<td>65</td>
<td>106</td>
<td>67</td>
<td>71</td>
<td>84</td>
<td>152</td>
<td>164</td>
<td>142</td>
<td>150</td>
<td>145</td>
<td>143</td>
</tr>
<tr>
<td>111</td>
<td>64</td>
<td>47</td>
<td>55</td>
<td>98</td>
<td>104</td>
<td>117</td>
<td>124</td>
<td>130</td>
<td>147</td>
<td>147</td>
</tr>
<tr>
<td>79</td>
<td>44</td>
<td>40</td>
<td>67</td>
<td>89</td>
<td>80</td>
<td>78</td>
<td>91</td>
<td>107</td>
<td>97</td>
<td>87</td>
</tr>
<tr>
<td>68</td>
<td>44</td>
<td>51</td>
<td>60</td>
<td>66</td>
<td>61</td>
<td>61</td>
<td>69</td>
<td>66</td>
<td>52</td>
<td>48</td>
</tr>
<tr>
<td>47</td>
<td>79</td>
<td>99</td>
<td>57</td>
<td>47</td>
<td>44</td>
<td>47</td>
<td>54</td>
<td>46</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>50</td>
<td>110</td>
<td>123</td>
<td>70</td>
<td>44</td>
<td>46</td>
<td>45</td>
<td>51</td>
<td>49</td>
<td>43</td>
<td>40</td>
</tr>
<tr>
<td>61</td>
<td>87</td>
<td>95</td>
<td>58</td>
<td>45</td>
<td>55</td>
<td>46</td>
<td>46</td>
<td>51</td>
<td>49</td>
<td>39</td>
</tr>
<tr>
<td>62</td>
<td>72</td>
<td>87</td>
<td>63</td>
<td>59</td>
<td>59</td>
<td>57</td>
<td>48</td>
<td>56</td>
<td>47</td>
<td>44</td>
</tr>
<tr>
<td>49</td>
<td>51</td>
<td>52</td>
<td>52</td>
<td>52</td>
<td>48</td>
<td>48</td>
<td>51</td>
<td>52</td>
<td>55</td>
<td>56</td>
</tr>
</tbody>
</table>
The image is focused onto the retina

Images are inverted by the eye

Why don’t we see everything upside down?

Perception can adapt to reversing the image:
SOME PRELIMINARY EXPERIMENTS ON VISION
WITHOUT INVERSION OF THE RETINAL IMAGE.

GEORGE M. STRATTON, 1896
“Seeing” the blind spot
The retina: An amazingly beautiful circuitry composed of many different cell types

- ~0.5 mm thick
- 5 x 5 cm retinal area
- Three cellular layers
- Rods (low-illumination conditions, ~$10^8$)
- Cones (color sensitivity, ~$10^6$)
- Blind spot
- Fovea (rod free, ~0.5 mm, ~1.7 deg)
- Midget ganglion cells (small dendritic arbors)
- Parasol ganglion cells (large dendritic arbors)

Dowling (2007), Scholarpedia, 2:3487
The retina: Some cells fire action potentials whereas other cells show graded responses

- Photoreceptors transduce incoming light input into electrical signals
- Rod to bipolar convergence increases rod-pathway sensitivity
- Cones, rods, horizontal and bipolar cells are non-spiking neurons
- Many different types of amacrine cells (some of which fire action potentials)
- Retinal ganglion cells fire action potentials and carry the output signals

Rods see largely in grayscale
There is much more detail at the fovea.
The retina has a huge dynamic range. Vision works well in moonless nights and in bright sunlight, a huge range of light intensity.

There are several mechanisms that support this adaptation:

- Changing pupil size
- Reliance on rods vs. cones
- Photopigment bleaching
- Feedback from horizontal cells to photoreceptors

These mechanisms are relatively slow.
The receptive field
Neurons throughout the visual system are very picky about the stimulus location.

This cartoon neuron responds only when a flash of light appears in the periphery, in the lower left quadrant.

Blumberg and Kreiman, 2010
Physiology of retinal ganglion cells

The receptive field of most RGCs have a center-surround structure.

Kuffler, S. (1953) J. Neurophys. 16: 37-68

About 1.2 million projections from each retina towards the brain.

Diversity of retinal ganglion cells

Minority of RGCs have more complex response properties:
- Phasic cells respond briefly to stimulus onset, offset, or both
- Some phasic cells respond selectively to edge orientation
- Suppressed-by-contrast cells fire except when an edge is present in receptive field
- Bistratified RGCs lack surrounds and are color-sensitive
- Color-opponent cells have centers and surrounds with opposing color preferences
- Intrinsically photosensitive RGCs contain photoreceptors and project to regions controlling pupil size, circadian rhythm, etc.
- Direction-sensitive cells respond to direction of motion of light or dark spots

These cells likely account for approximately 10% of RGCs
Unclear to what extent, these cells contribute to visual object recognition

Stone and Fukuda, *Journal of Neurophysiology* 1974
Cleland and Levick, *Journal of Neurophysiology* 1974
Berson et al., *Science* 2002
The lateral geniculate nucleus (LGN) is the main visual part of the thalamus:
• 6 layers, contralateral visual hemifield
• Layers 2, 3 and 5 receive ipsilateral eye’s input
• Layers 1, 4 and 6 receive contralateral eye’s input
• Layers 1-2: magnocellular layers that receive input from parasol ganglion cells
• Layers 3-6: parvocellular layers that receive input from midget ganglion cells
• Between the layers: koniocellular layers that receive input from bistratified retinal ganglion cells
• Right and left visual hemifields are separate in the LGN
• Right and left eyes are separate in the LGN
• The visual field is represented multiple times in the LGN
• On and Off center cells are present in all layers
• LGN does not project back to the retina

NOTE: Most of the input to the LGN comes from visual cortex and not from the retina! (e.g. Douglas and Martin 2004)

Difference of Gaussians

The center-surround structure can be described by a difference of gaussians (Mexican-hat)

\[
D(x,y) = \pm \left( \frac{1}{2\pi\sigma_{cen}^2} \exp\left[-\frac{x^2 + y^2}{2\sigma_{cen}^2}\right] - \frac{B}{2\pi\sigma_{sur}^2} \exp\left[-\frac{x^2 + y^2}{2\sigma_{sur}^2}\right] \right)
\]

Neurons respond with transient bursts of activity

Dynamic receptive fields in the retina/LGN

\[
D(x,y,t) = \pm \left( \frac{D_{\text{cen}}(t)}{2\pi\sigma_{\text{cen}}^2} \exp\left[-\frac{x^2 + y^2}{2\sigma_{\text{cen}}^2}\right] - \frac{BD_{\text{sur}}(t)}{2\pi\sigma_{\text{sur}}^2} \exp\left[-\frac{x^2 + y^2}{2\sigma_{\text{sur}}^2}\right] \right) 
\]

\[
D_{\text{cen}}(t) = \alpha_{\text{cen}}^2 t \exp[-\alpha_{\text{cen}} t] - \beta_{\text{cen}}^2 t \exp[-\beta_{\text{cen}} t] 
\]

\[
D_{\text{sur}}(t) = \alpha_{\text{sur}}^2 t \exp[-\alpha_{\text{sur}} t] - \beta_{\text{sur}}^2 t \exp[-\beta_{\text{sur}} t] 
\]

Difference of Gaussians in space and time
The center-surround structure can also be seen in receptive field dynamics

Subcortical visual pathways

Retinal projections

Lateral geniculate nucleus (LGN) – Thalamus

Superior Colliculi – Main visual pathway in birds, reptiles, fish

Implicated in saccade generation in mammals

Suprachiasmatic Nucleus – Hypothalamus: involved in circadian rhythms

Pretectum

Pregeniculate

Accessory optic system

Primates can recognize objects after lesions to the Superior Colliculus but not after lesions to V1 (Gross 1994 for historical overview).
Visual system circuitry

You are here

Felleman and Van Essen. Cerebral Cortex 1991
Further reading

• Class notes: http://tinyurl.com/vision-class

Some of the original articles cited in class (see lecture notes for full list)

• Simoncelli and Olshausen. Annual Review of Neuroscience 2001
• Dowling J. Scholarpedia 2007.
• Felleman and Van Essen. Cerebral Cortex 1991.
• Kuffler. Journal of Neurophysiology 1953.
Cited works


