Lecture 6: Psychophysical studies of visual object recognition

We can learn a lot about visual object recognition by carefully quantifying performance under a variety of well-controlled visual tasks. A discipline with the peculiar and attractive name of “Psychophysics” aims to rigorously characterize and constrain the behavior during cognitive tasks.

6.1 What you get ain’t what you see

It is clear that what we end up perceiving is a significantly transformed version of the pattern of photons impinging on the retina. Let us start with some very simple examples. The Gestalt laws provide basic constraints about how patterns of light are integrated into perceptual sensations (Reagan, 2000).

- **Law of closure.** We complete lines and extrapolate to complete known patterns or regular figures. An example of this is given by the famous Kanizsa triangle. Our mind creates a triangle in the middle of the image from incomplete information (Figure 6.1).

- **Law of similarity.** We tend to group similar objects together. Similarity could be defined by shape, color, size or brightness (Figure 6.2)

- **Law of proximity.** We tend to group objects based on their distance (Figure 6.3).

- **Law of symmetry.** We tend to group symmetrical images.

- **Law of continuity.** We tend to continue regular patterns (Figure 6.4).

- **Law of common fate.** Elements with the same moving direction tend to be grouped.

These laws are usually summarized by pointing out that the forms (Gestalten) are more than the mere sum of the component parts.

Figure 6.1. The Kanizsa triangle. The mind creates a white triangle from the incomplete information provided by the pacman or other shapes in the figure.

Figure 6.2. Law of similarity. We tend to group objects that share some properties.
6.2 Tolerance to object transformation

A hallmark of visual recognition is our ability to identify and categorize objects in spite of large transformations in the image. An object can cast an infinite number of projections onto the retina due to changes in position, scale, rotation, illumination, color, etc.

You can easily demonstrate the strong degree of tolerance for object transformations. For example, take a piece of text with 12pt font size, hold it at arm’s length and focus on any given letter, say “A”. The A will subtend a fraction of one degree of visual angle (approximately the size of your thumb at arm’s length).

A particularly intriguing example of tolerance is given by the capability to recognize caricatures and line drawings. At the pixel level, these images seem to bear little resemblance to the actual objects and yet, we can recognize them quite efficiently, sometimes even better than the real images!

6.3 Speed of visual recognition

Visual recognition seems almost instantaneous. Several investigators have shown that we can recognize complex objects in a small fraction of a second.

One of the original studies by Mary Potter consisted of showing a sequence of images in a rapid sequence (RSVP, rapid serial visual presentation) and showing that subjects could detect the individual images even when presented at rates of 8 per second (Potter and Levy, 1969). Complex objects can be recognized when presented tachistoscopically for < 50 ms without a mask, even in the absence of any prior expectation or other knowledge (Vernon 1954).

Part of the delays in reaction time measurements are associated with the behavioral response. In an attempt to constrain the amount of time required for visual recognition, Thorpe and colleagues recorded evoked response potentials from scalp electroencephalographic (EEG) signals while subjects performed a go/no-go animal categorization task (Thorpe et al., 1996). They found that frontal cortex electrodes showed a differential signal at about 150 ms; they argued that visual discrimination of animals versus non-animals in complex scenes should happen before that time. Kirchner et al used eye movements to elicit rapid responses and showed that subjects could make a saccade to discriminate the presence of a face or non-face stimulus in slightly more than 100 ms (Kirchner...
and Thorpe, 2006). These observations place a strong constraint into the mechanisms that underlie visual recognition.

6.4 Beyond pixels – context and other effects

Several visual illusions demonstrate the existence of strong contextual effects in visual object recognition. For example, it is significantly more difficult to recognize faces when they are upside down. In a simple yet elegant demonstration, the perceived size of a circle can be strongly influenced by the size of its neighbors (Figure 6.5). Several entertaining examples of contextual effects have been reported (e.g. (Eagleman, 2001; Sinha and Poggio, 1996)).

Figure 6.5. Context matters. The green circle on the right appears larger than the one on the left but they are the same size.

6.5 The value of experience

Our percepts are also influenced by previous visual experience. This observation holds at multiple different temporal scales. At short time scales, several visual illusions show the powerful effects of visual adaptation. One such illusion is the waterfall effect: after staring at a waterfall for a minute or so, and then shifting the gaze to other static objects, those objects appear to be moving upward. At longer time scales, the interpretation of an image could depend on whether one has seen the image before. A typical example is the Dalmatian dog: for the first-time observer the image consists of a smudge of black and white spots. However, after recognizing the dog, people can immediately spot him the next time. Other similar examples are Mooney images (Mooney, 1957).

References