BEWARE: These are preliminary notes. In the future, they will become part of a textbook on Visual Object Recognition.

Chapter III. Psychophysical studies of visual object recognition

We want to understand the neural mechanisms responsible for visual object recognition and we want to instantiate these mechanisms into computational algorithms that resemble and perhaps eventually surpass human performance. In order to untangle the mechanisms orchestrating visual recognition and build adequate computational models, we need to better define the problem visual recognition capabilities at the behavioral level. What shapes can humans recognize and when and how? Under what conditions do humans make mistakes? How fast can humans recognize complex objects? We can learn about visual object recognition by carefully quantifying human performance under a variety of well-controlled visual tasks. A discipline with the peculiar and attractive name of "Psychophysics" aims to rigorously characterize, quantify and understand behavior during cognitive tasks.

3.1. 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. Our brains filter and process visual inputs to understand the physical world by constructing an interpretation that is consistent with our experiences.

A simple example of the dissociation between inputs and percepts is given by the blind spot. If you close one eye, there is a part of the visual field that is not mapped onto retinal ganglion cells, the spot where these cells leave the retina to form the optic nerve. It is possible to distinguish this blind spot by closing one eye, fixating on a given spot and slowly moving a finger from the center to the periphery until part of it disappears from view (but not in its entirety which would imply that you moved your finger completely outside of your visual field).

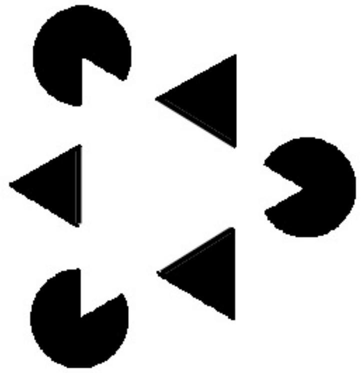

Figure 3.1: The Kanizsa triangle. The mind creates a white triangle from the incomplete information provided by the pacmen or other shapes in the figure.

3.2. Gestalt laws of grouping

The *Gestalt* laws (in German "gestalt" means shape) provide basic constraints about how patterns of light are integrated into perceptual sensations

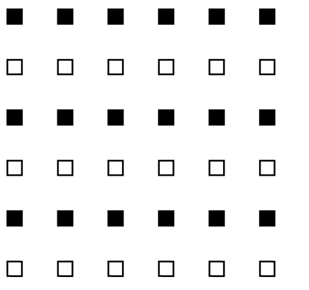

Figure 3.2. Law of similarity. We tend to group objects that share some properties.

(Reagan, 2000). These rules arose from attempts to understand the basic perceptual principles that lead to interpreting objects as wholes rather than the constituent isolated lines or elements that give rise to them.

- *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 3.1**).
- *Law of similarity*. We tend to group similar objects together. Similarity could be defined by shape, color, size or brightness (**Figure 3.2**)
- *Law of proximity*. We tend to group objects based on their distance (**Figure 3.3**).
- Law of *symmetry*. We tend to group symmetrical images.
- *Law of continuity*. We tend to continue regular patterns (**Figure 3.4**).
- *Law of common fate*. Elements with the same moving direction tend to be grouped.

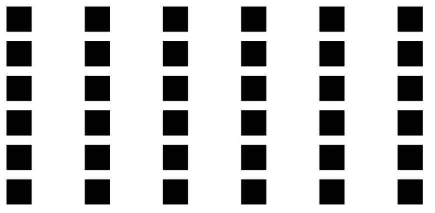

Figure 3.3. Law of proximity. We tend to see this figure as vertical lines.

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

3.3. Holistic processing

Some of the most compelling examples of the processing and interpretation of a whole image beyond what can be discerned from the individual components is the example of holistic processing of faces. Three main observations have been put forward to document the holism of face processing. First is the inversion effect (Valentine, 1988; Yin, 1969), which describes how difficult it can be to distinguish local changes in a face when it is turned upside down (this is also called the "Thatcher effect" alluding to the images of Britain's prime minister used to demonstrate the perceptual illusion). The second observation is the composite face illusion: by putting together the upper part of face 1 and

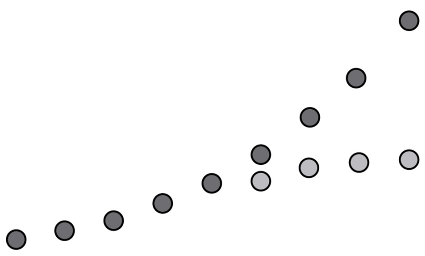

Figure 3.4. Law of continuity. We tend to assume that the dark gray circles form a line.

the bottom part of face 2, one can create a novel face that appears to be perceptually distinct from the two original ones (Young et al., 1987). A third argument for holistic processing is the parts and wholes effect: changing a local aspect of a face distorts the overall perception of the entire face (Tanaka and Farah, 1993).

3.4. Tolerance to object transformations

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. Our visual system shows a significant degree of robustness to changes in:

- Scale (e.g. you can recognize an object at different sizes). 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).
- Position with respect to fixation (e.g. we can recognize an object placed at different distances to the fixation point)
- 2D rotation (e.g. we can recognize an object after turning our head sideways or rotating the object within the plane)
- 3D rotation (e.g. we can recognize an object from different viewpoints)
- Color (e.g. we can recognize the objects in a photograph whether it's in color, sepia, grayscale)
- Illumination (e.g. consider illuminating an object from the left, right, top or bottom)
- Cues (e.g. an object's shape can be determined by edges, by motion cues, by completion without sharp edges)
- Clutter (e.g. we can recognize objects despite the presence of other objects in the image)
- Occlusion (e.g. we can recognize objects from partial information)
- Other non-rigid transformations (e.g. we can recognize faces even with changes in expression, aging, even from the line drawing sketches in **Figure 3.5**!)

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!

3.5. Speed of visual recognition

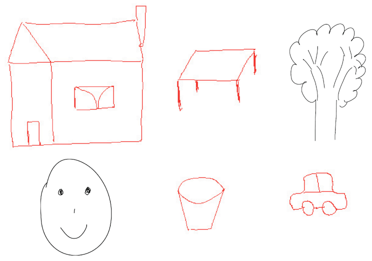

Figure 3.5. Recognition of line drawings. We can identify the objects in these line drawings despite the extreme simplicity in the traces and the minimal degree of resemblance to the actual objects..

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.

3.6. Beyond pixels – contextual 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 (see “Holistic processing” above). In a simple yet elegant demonstration, the perceived size of a circle can be strongly influenced by the size of its neighbors (**Figure 3.6**).

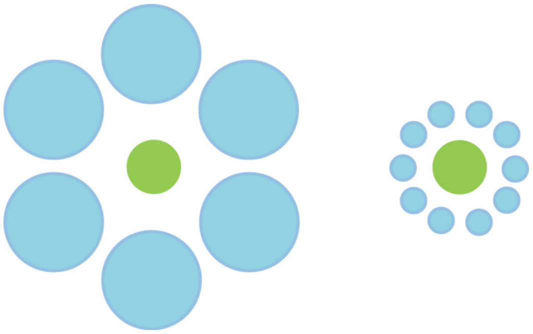

Figure 3.6. Context matters. The green circle on the right appears larger than the one on the left but they are the same size.

Another extremely simple example is the Muller-Lyer illusion: the perceived length of a line with arrows at the two ends depends on the directions of the two arrows. Several entertaining examples of contextual effects have been reported (e.g. (Eagleman, 2001; Sinha and Poggio, 1996)). These strong contextual dependences illustrate that the visual system spatially integrates information and the perception of local features depends on the more global surrounding properties. The contextual effects also emphasize that perception constitutes an interpretation of the input in the light of context and experience.

4.5 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

Eagleman, D.M. (2001). Visual illusions and neurobiology. Nat Rev Neurosci 2, 920-926.
Kirchner, H., and Thorpe, S.J. (2006). Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. Vision research 46, 1762-1776.
Mooney, C.M. (1957). Age in the development of closure ability in children. Can J Psychol 11, 219-226.
Potter, M., and Levy, E. (1969). Recognition memory for a rapid sequence of pictures. Journal of experimental psychology 81, 10-15.
Reagan, D. (2000). Human perception of objects (Sinauer).
Sinha, P., and Poggio, T. (1996). I think I know that face. Nature 384, 404.
Tanaka, J.W., and Farah, M.J. (1993). Parts and wholes in face recognition. The Quarterly journal of experimental psychology A, Human experimental psychology 46, 225-245.
Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. Nature 381, 520-522.
Valentine, T. (1988). Upside-down faces: a review of the effect of inversion upon face recognition. Br J Psychol 79 (Pt 4), 471-491.
Vernon, M. (1954). Visual perception (Cambridge: Harvard University Press).
Yin, R. (1969). Looking at upside-down faces. Journal of experimental psychology 81, 141-145.
Young, A.W., Hellawell, D., and Hay, D.C. (1987). Configurational information in face perception. Perception 16, 747-759.