CHAPTER 16

OBJECT PERCEPTION AND SCENE ANALYSIS

16.1 SCENE ANALYSIS

Scene analysis is the process through which all of the information taken in through one or more sensory modalities produces a representation and awareness of the environment. It depends on capturing, processing, and organizing information. In vision, scene analysis results in perception of objects, their significance, and their relationships.

16.1.1. Transformation from stimulation of the retinal array to perceived objects.

The retinal image can be thought of as a 2-dimensional array of points of illumination. Beginning at the retina, our nervous system processes and organizes this information so that we do not perceive points of light, but instead perceive 3-dimensional objects.

16.1.2. Features that are important for object perception.

**Changes in illumination over time.** Changes in illumination are necessary in order for us to see anything at all. Changes in the pattern of illumination across the retina are provided by movement of objects in the environment and by movement of the eyes relative to the environment. There are several ways in which the eyes may move relative to the environment. The largest movements occur whenever the head and/or body moves relative to the visual field. In addition, the eyes themselves move around so that the fovea is directed to different portions of the visual field. These scanning movements of the eyes relative to contours in the visual field are called *saccades*: In addition to these large-scale scanning movements of the eyes, there are constant tiny movements or "jiggling" of the eyes that occur even when you think your eyes are held perfectly still. These small jiggling movements are called microsaccades. Visual neurons respond best to a *change* in illumination, regardless of whether it is an increase, a decrease, or a qualitative (color) change. Microsaccades ensure that there is always change across the array of receptors on the retina.

**Contours.** Contours, or local variations in the intensity or quality of illumination, are the fundamental elements of visual perception. Without contours, we could not see anything meaningful.

**Figure-ground relationships.** A figure is an integrated group of contours or perceptual elements; the ground is the surface (background) against which the figure appears. Generally, smaller areas are seen as figures, larger ones as ground. It is impossible to perceive an area as both figure and ground simultaneously, although it is possible to make a rapid cognitive switch back and forth between figure and ground when viewing an ambiguous scene.

**Subjective (illusory) contours.** Illusory contours are not physically present on the retina, but are nevertheless part of our perception of a figure/ground relationship. Our brain "fills in" the missing contours, presumably because we expect to see familiar patterns that we have learned through their frequent occurrence in the environment.
**Figural grouping.** Most forms or objects that we see are composed of many different elements. These elements are organized into groups that are perceived as belonging together.

![Figural grouping examples](image)

**Figure 16-1.** Ambiguous figure/ground scenes. In the panel on the left, the white area is smaller and tends to be perceived as a figure against a black background. In the panel on the right, the black areas are smaller and tend to be seen as two black figures against a white background.

![Illusory contours examples](image)

**Figure 16-2.** Illusory contours are seen when there are contours that suggest the presence of other contours corresponding to a familiar form. In the illustration on the left, we perceive letters covered by a white bar. In the illustration on the right, we perceive a white triangle partially covering another white triangle, the corners of which partially cover three dark squares.

### 16.2. THEORIES OF OBJECT PERCEPTION

There are many different theories of how our brains take the raw information that exists at the photoreceptor array and convert it to a perceptual experience that includes objects and their spatial and temporal relationships. This section provides some examples of theories that have been proposed to explain object perception.

#### 16.2.1. Gestalt theory.

*Gestalt* is a German word that literally means "form". Gestalt theory, first proposed in the 19th century, attempts to explain (in non-neurobiological terms) how sensory systems organize information so that we are able to perceive objects or forms. The underlying principle, "The whole is greater than the sum of its parts"...summarizes the Gestalt approach to perceptual
grouping and understanding object perception. These same principles can also be applied to explaining perceptual organization in sensory systems other than vision.

16.2.1.1. The "Gestalt Laws" or principles of grouping

**Prägnanz** (a.k.a. law of "good figure", or law of simplicity).
For a given group of stimulus elements, the form perceived is the simplest one possible.

![Figure 16-3](image)

*Figure 16-3.* Different arrangements of lines that vary in the probability that they will be perceived as a complex 3-dimensional object, even though any of them theoretically could be a box viewed at a specific angle. The tendency is to see the two objects on the left as flat and the two on the right as having depth.

**Similarity.** In an array of different stimulus elements, similar elements are grouped together.

![Figure 16-4](image)

*Figure 16-4.* When there are multiple elements in an array, we tend to group similar elements (e.g., circles, squares) together to form a figure embedded in a background or to form separate figures.

**Proximity.** In an array of stimulus elements, those that are close to one another tend to be grouped together.
**Figure 16-5.** In an array of identical elements, ones that are close together tend to be lumped into a coherent group that may appear as a shape (circles on left) or groups of elements that "belong together" (pairs of lines on right).

*Good continuation.* Stimulus elements that appear to follow in the same direction tend to be grouped together.

![Figure 16-5](image)

**Figure 16-6.** Left: The dots that follow one another in a spiral pattern are grouped together, but the "odd" dot at the top is left out of the pattern. Right: This pattern may be organized in a number of different ways depending on whether the eye focuses on the larger or smaller patterns.

*Familiarity.* Stimulus elements that form a familiar shape tend to be grouped together.

![Figure 16-6](image)

**Figure 16-7.** Familiarity with specific stimuli in specific contexts causes us to perceive the most familiar grouping of elements. For example, a pair of shapes are grouped together to form a "B" when they appear in a series of letters. The identical shapes are viewed as separate, forming the number 13, when viewed in a series of numbers.

*Common fate.* Stimulus elements that move together tend to be grouped together.

*Closure.* A space enclosed by a contour (or partial contour) tends to be perceived as a figure.
16.3. FEATURE EXTRACTION THEORIES

In Gestalt theory, there are no assumptions about stages of processing. In fact, it is assumed that all aspects of the sensory scene are processed together or in parallel. In fact, the prevalence of parallel pathways and feedback loops in the brain provides biological evidence to support this view. Other, more recent theories incorporate the known hierarchical structure of brain pathways into various conceptual views of how forms and objects are perceived. Of course, the theories are not mutually exclusive. It is very likely that sensory processing includes aspects of all of the theories.

16.3.1. Feature integration theory

According to feature integration theory, there are 2 major stages in the perception of forms/objects. These are:

**Preattentive processes** (i.e., processes that happen without active attention) extract contours and fundamental features such as color, size, line orientation, representing the most basic components present in the visual field. These fundamental features have been given different names including "primitives" and "textons", but the basic idea is the same. The extraction of these fundamental “features” would roughly correspond to the progressive development of specialized neural tuning up through the level of primary visual cortex. At some intermediate, “pre-conscious” stage, the contours (primitives) are grouped into preliminary figures.

**Focal attention** refers to the process whereby the observer's attention is focused on a particular location within the visual field so that primitives and/or preliminary figures are grouped into perceived objects, sometimes referred to by the jargon term “object files”. The object files are then compared with some form of memory template to identify the object.

16.3.2. Recognition-by-Components

According to this theory, the process that occurs is similar to that postulated by the feature integration theory except that instead of contours and colors, the primitives are simple 3-D geometric forms, or "geons" which are then grouped to form identifiable objects.
16.3.3. Spatial frequency theory

Any sensory stimulus that contains a pattern of change across a dimension can be broken down (by Fourier analysis) into a series of sine-wave components, defined in the relevant dimension - space, in the case of a visual scene. In other words, any visual stimulus can be broken down into a series of sine-wave gratings, each of which has a "spatial frequency". When added together, this series of gratings produces the 2-dimensional pattern of illumination present in the visual scene.

Regardless of whether it exists in the temporal or spatial dimension, or some other dimension, a waveform such as a sine wave has a frequency (cycles/unit of space or time) and an amplitude (intensity). A visual sine-wave stimulus is characterized by its frequency (cycles per unit distance) and its contrast (amount by which intensity is modulated around a center value).

16.3.3.1. Frequency channels in the nervous system. Theoretically, the visual system could have different neural "channels", each of which is sensitive to a different range of spatial frequencies. Together these channels could break a complex pattern down into its frequency components (i.e., perform a Fourier analysis). We are most sensitive to spatial frequencies of 2-4 cycles/degree, at the fovea, so will be most likely to pick out details of patterns having this spatial frequency.

Thought question: Given what you know about the neural basis of information processing in the visual system, why do you think we might "fill in" missing contours such as the sides of the triangle in Figure 16-6?

Figure 16-9. A scene made up of areas with different spatial frequencies. Note that the contrast between black and white is abrupt, so this pattern of transition would correspond to a square wave grating, not a sine wave grating where the transition from black to white and back again is gradual.
Thought question: In the scene shown in Figure 16-9, which area(s) are we most likely to perceive as "figure" and which as "background"? Why? What would we see if the spatial frequencies were reversed?

Thought question: Feature extraction theory assumes that analysis of a visual scene occurs in stages with analysis of "simple" features performed prior to analysis of more complex features. Gestalt theory makes no such assumption - in fact, all processing could be parallel. What evidence is there to support each position? Suggest an experiment using human subjects that would distinguish between feature extraction in stages vs parallel processing of a visual scene.

Thought question: Assuming that different populations of neurons are tuned to different ranges of spatial frequency, what neural processing mechanisms might underlie the differential tuning?