Computer Vision (CE316 and CE866): The Human Visual System

Adrian F. Clark
⟨*alien@essex.ac.uk*⟩

Before considering computer vision in any detail, it is instructive to look at some of the capabilities of the human visual system. We shall start with the eye itself, which is fairly well understood, and then move on to what we know of the processing performed by the brain. Note that the content of this chapter is to aid your understanding; it is not examinable.

Contents

List of Figures

1 The eye

Optically, there are close analogies between a camera and the eye. The front of the eye is covered by the *cornea*, which performs about 80% of the focusing that is needed. Behind that is the *lens*, whose shape and hence optical power can be varied a little by ciliary muscles, in much the same way that a camera can be focused. This focusing ability is known as *accommodation*. A young individual with no optical defects can focus easily down to a *near point* of about 10 cm, though as this individual ages, accommodation degrades and the near point moves further away — this is known as becoming *long-sighted* and its medical term is *presbyopia*. People who cannot focus on infinity — a common trait among people who have done a lot of close work when young — are *short-sighted* (the medical term is *myopia*) and generally wear spectacles or contact lenses all the time.

Around the lens lies the *iris*, which controls the area of the lens through which light is able to enter, commonly known as the *pupil*. Pupil sizes generally lie in the range 3–7 mm and correspond to about a five times change in pupil area and hence the amount of light allowed into the eye.

The back of the eye, the *retina* (Figure 1), is covered with a layer of light-sensitive cells that is about the thickness of tissue paper. There are two main types of cells, known as *rods* and *cones*. The rods are responsible for vision in low light ($< 1\,\mathrm{cd/m^2}$), while the cones require more light to work. It is generally believed that there are three types of cones, having pigments that are more responsive in the red, green and blue parts of the spectrum (Figure 2), and this is what gives us colour vision. The red-, green- and blue-sensitive cones occur roughly in the ratio 12 : 6 : 1, so the eye is significantly less sensitive to blue light than to green or red, as Figure 2 shows. It is postulated that this is because humans evolved in a region where the staple diet was red berries growing amongst green foliage, though there is no way of verifying this. The presence of three different types of cone explains why some people are colour-blind: *e.g.,* there is a minor defect on the male chromosome that affects the pigmentation of the red- and green-sensitive cones in about 10% of men.

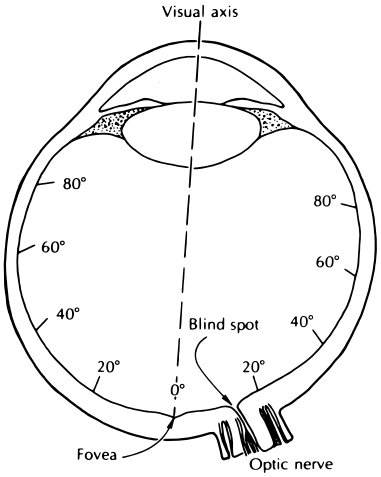


Figure 1: A cross-section through the human eye (from [Cornsweet, 1970] p137). The angles marked around the retina correspond to the angles in Figure 4.

The rods are more sensitive in the green–blue part of spectrum than cones (Figure 3), presumably because of the need to retain some vision functionality in moonlight, which is bluish in colour. This also explains why ships normally illuminate their bridges with red light at night: it avoids destroying the dark-adaption of the rods.

The eye contains about 50,000 cones and about 120 million rods, about twenty times more — though groups of about 120 rods are 'wired together' into a single ganglion cell (see below) to amplify the faint responses that result from low illumination, while only about 6 cones converge into one ganglion. The distribution of rods and cones over the eye is far from uniform. Both rods and cones are roughly circular in cross-section and hence tend to pack together somewhat more like hexagons then rectangles — unlike, say, the sensors on most CCD cameras. In humans and most other predators, the cones are concentrated into a region known as the *fovea centralis* (Figure 1 and 4), which is almost rod-free.

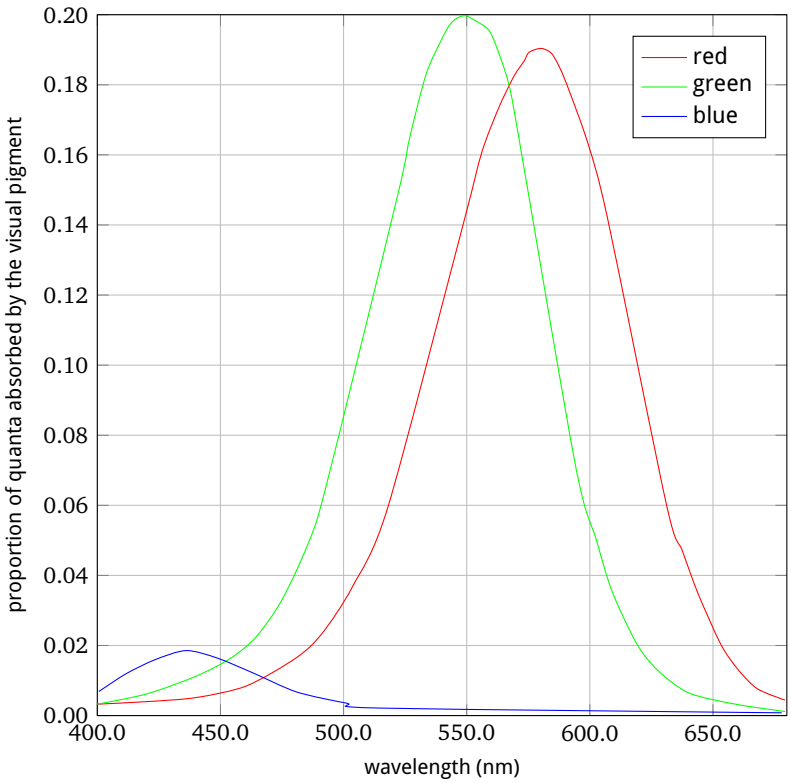


Figure 2: The relative spectral sensitivity of cones (adapted from [Cornsweet, 1970] p171, where it is reported that the subject from which these measurements were taken was about three times more blue-sensitive than an average subject).

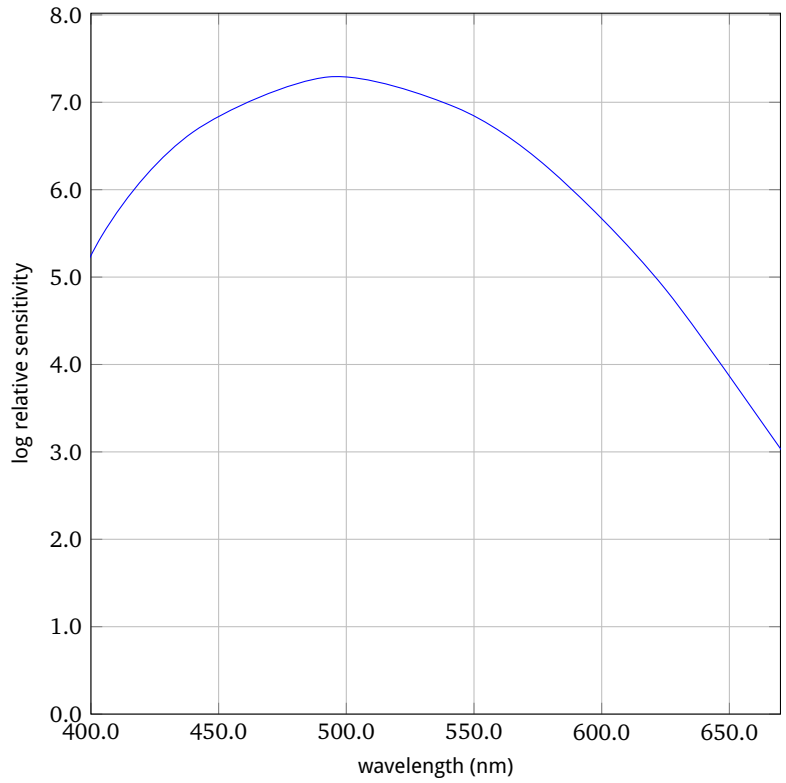


Figure 3: The spectral sensitivity of rods (adapted from [Cornsweet, 1970] p146)

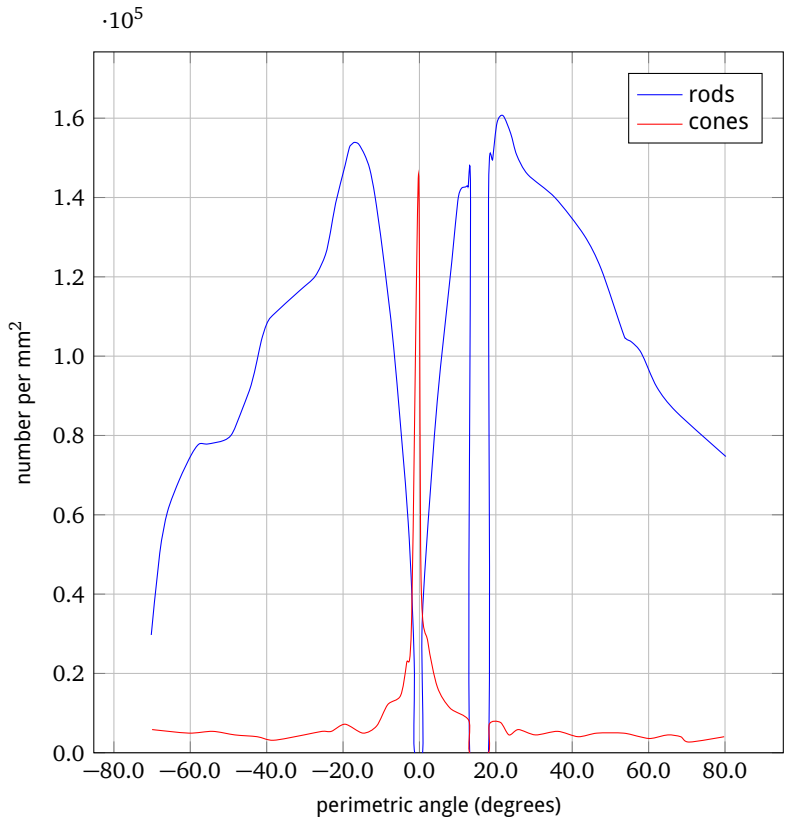


Figure 4: The distribution of rods and cones on the retina (adapted from [Cornsweet, 1970] p137). The angles correspond to those marked on Figure 1.

(Some birds, for example, have more than one fovea.) In humans, the fovea subtends about 2° of visual angle and contains about 1% of the cones. A higher concentration of sensors means that the resolving power is higher, *i.e.* that greater detail can be perceived, and that colour discrimination is better. Without realising it, people turn their heads and eyes so that incoming light is focused on the fovea and hence the perception of detail (and colour) is greatest (*fixation*). The better light-gathering efficiency of the rods around the periphery of the fovea is the reason that star-gazers are recommended to look 'above' stars rather than directly at them. Moreover, when one goes into the dark, it takes cones about six minutes to become maximally sensitive and rods about 30 minutes — and this is one reason that star-gazing is more effective in places without street lights and well away from roads.

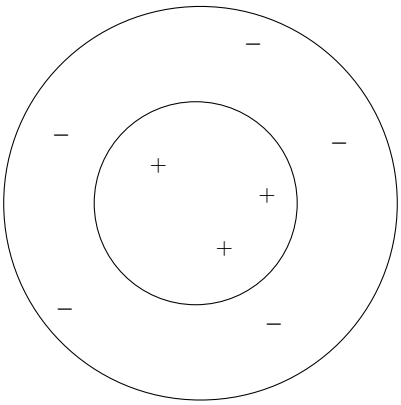
2 From the eye to the brain

The electrical signals generated by photons hitting the rods and cones flow through four different types of cells (*amacrine*, *bipolar*, *horizontal* and *ganglion* cells), and the axons of the latter form the *optic nerve* which runs from the eye into the brain. The place where the optic nerve leaves the eye contains no rods or cones and hence leaves a blind spot (Figure 4). Fortunately, this is not apparent to us, partly because of fixation and partly because the brain appears to interpolate over the blind spot.

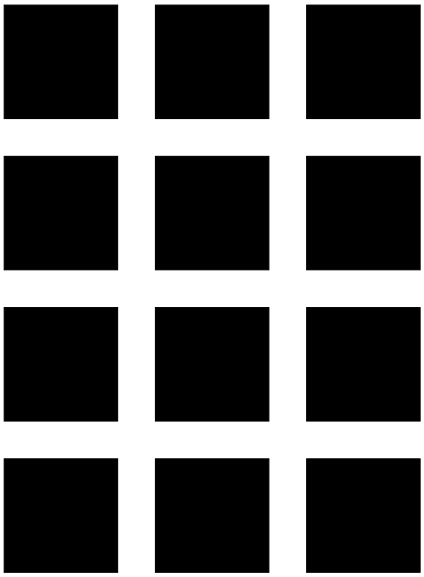
The way in which the signals from the retina converge into a ganglion

cell provides an excitatory centre and inhibitory surround (Figure 5(a)) which, as we shall see when we consider convolution, is similar to some image operators. This phenomenon explains the *Hermann grid* (Figure 5(b)) in which grey 'ghosts' appear in the intersections of white 'corridors,' and Mach banding, an illusion that affects shaded objects in computer graphics near to illumination edges. The same effect causes *simultaneous contrast*, in which the apparent brightness of an area is affected by the brightness of its surroundings, and is thought to help *continuity*, where shapes that have incomplete boundaries are 'filled in' by the brain.

The optic nerves travel principally from the eyes to regions of the brain known as the *lateral geniculate nucleus* (LGN), though about 10% go to the *superior colliculus*, which controls eye movement. The LGN also receives signals from other parts of the brain, such as the *visual cortex* and the *thalamus*. There is one LGN in each half of the brain, and each of them comprises six distinct layers. Each eye sends half of its signals to the left LGN and the other half to the right LGN. The layers of the LGN retain the relationship between rods and cones on the retina, so we can think of it as working on images.



(a) The excitatory centre is surrounded by an inhibitory region



(b) One consequence of this is the Hermann grid, where grey spots are apparent at the junctions of the white lines

Figure 5: The excitatory–inhibitory nature of ganglion cells and its consequences

3 Inside the brain

About 1.5 million axons travel from the LGN to the region of the *visual striate cortex* known as *V1* (again, there is one in each half of the brain). The cortex contains about 250 million neurons and is responsible for most of the higher-level processing in the human visual system. Cells in the V1 region respond to specific aspects of images, such as orientation, edge and motion (Hubel and Wiesel won the Nobel prize for identifying this in 1950); for this reason, cells in the striate cortex are often known as *feature detectors*, and this is one of the reasons that vision researchers have expended a great deal of energy in developing edge, corner and feature detectors. Clearly, just like the LGN, cells in this region must relate to specific points in the retina. Interestingly, 8–10% of the cells are connected to the fovea even though it corresponds to about 0.01% of the number of cells on the retina. Each V1 transmits information to two pathways:

The dorsal stream goes through visual area V2, on to V5 and thence to the posterior parietal cortex. This stream is sometimes called the "where pathway" or "how pathway" as it is associated with motion, the representation of object locations, and control of the eyes and arms.

The ventral stream also goes through visual area V2 then through visual area V4 and on to the inferior temporal cortex. This stream is sometimes called the "what pathway" as it is associated with form recognition, object representation and the storage of long-term memory.

The four regions of V2 appear to produce a map of the visual world. As in V1, cells are tuned to properties of the visual field such as orientation, spatial frequency, and colour; however, the responses of many V2 neurons are also modulated by more complex properties, such as the orientation of illusory contours and whether the stimulus is part of foreground or background.

Beyond V2, detailed knowledge of the processing is less certain. It is thought that V3 plays a rôle in the determining global motion in the visual signal while V4, like V1, is tuned for orientation, spatial frequency, and colour. However, V4 is tuned for more complex object features than V1, for example simple geometric shapes (but not faces). V5 is a region of *extrastriate visual cortex* which is thought to play a major rôle in the perception of motion, the integration of local motion signals into global percepts and the guidance of some eye movements.

There is some knowledge of where the processing of particular types of features takes place in the brain, usually acquired from studies of people who have suffered oxygen starvation in strokes or road traffic accidents, resulting in the death of specific brain regions. For example, the ability to store and recognise faces appears to be centralised in one region of the brain, and similarly the conversion of 2D knowledge of shapes into 3D recognition appears to take place in a specific region. There are theories as to some of the mechanisms involved but it is fair to say that we are far from a complete understanding — certainly too far for us to develop computer vision systems simply by simulating what happens in the brain. There are a few Nobel prizes to be won before that becomes a realistic possibility.

References

Cornsweet, Tom N. (1970). *Visual Perception*. Academic Press.