Chapter 4

Human Vision

4.1 The Visual System

The human visual system can be regarded as consisting of two parts. The eyes act as image receptors which capture light and convert it into signals which are then transmitted to image processing centres in the brain. These centres process the signals received from the eyes and build an internal "picture" of the scene being viewed. Processing by the brain consists of partly of simple image processing and partly of higher functions which build and manipulate an internal model of the outside world.

Although the division of function between the eyes and the brain is not clear-cut, it is useful to consider each of the components separately.

4.2 The Eye

The structure of the human eye is analogous to that of a camera. The basic structure of the eye is displayed in figure 4.1

- The *cornea* and *aqueous humour* act as a primary lens which perform crude focusing of the incoming light signal.

- A muscle called the *zonula* controls both the shape and positioning (forward and backwards) of the eye's *lens*. This provides a fine control over how the light entering the eye is focused.

- The *iris* is a muscle which, when contracted, covers all but a small central portion of the lense. This allows dynamic control of the amount of light entering the eye, so that the eye can work well in a wide range of viewing conditions, from dim to very bright light. The portion of the lens not covered by the iris is called the *pupil*.

- The retina provides a photo-sensitive screen at the back of the eye, which incoming light is focused onto. Light hitting the retina is converted into nerve signals.

- A small central region of the retina, called the *fovea*, is particularly sensitive because it is tightly packed with photo-sensitive cells. It provides very good resolution and is used for close inspection of objects in the visual field.

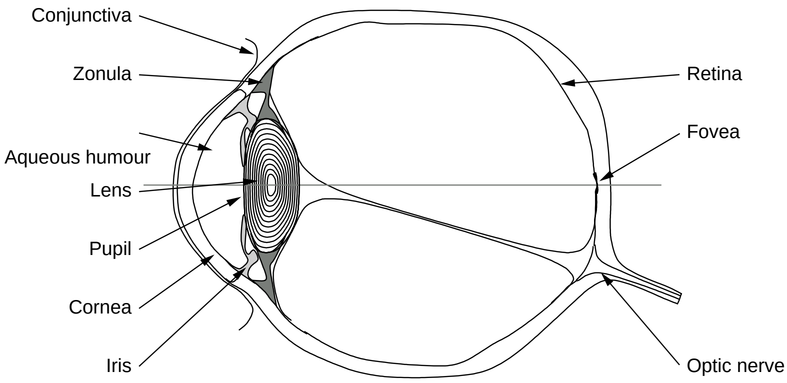

Figure 4.1: A cross-section of the right human eye, viewed from above.

- The optic nerve transmits the signals generated by the retina to the vision processing centres of the brain.

4.2.1 The Blind Spot

The area of the retina where the optic nerve is attached is completely devoid of photosensitive cells. This means that there is a “blind spot” in the field of vision for each eye. Most of the time we are not aware of this deficit in our vision, but it is quite easy to locate it. Close your right eye and stare that the cross in the figure below with your left eye. Keep staring at the cross and move the page closer to your eye. At some point the dot in figure will disappear.

If you move the page even closer, and the dot will reappear.

4.2.2 The Retina

The retina is composed of a thin layer of cells lining the interior back and sides of the eye. Many of the cells making up the retina are specialised nerve cells which are quite similar to the tissue of the brain. Other cells are light-sensitive and convert incoming light into nerve signals which are transmitted by the other retinal cells to the optic nerve and from there to the brain.

There are two general classes of light sensitive cells in the brain; *rods* and *cones*. Rod cells are very sensitive and provide visual capability at very low light levels. Cone cells perform best at normal light levels. The provide our daytime visual facilities, including the ability to see in colour (which we discuss in the next chapter).

There are roughly 120 million rod cells and 6 million cone cells in the retina. There are many more rods than cones because they are used at low light levels and so more of them are required to gather the light.

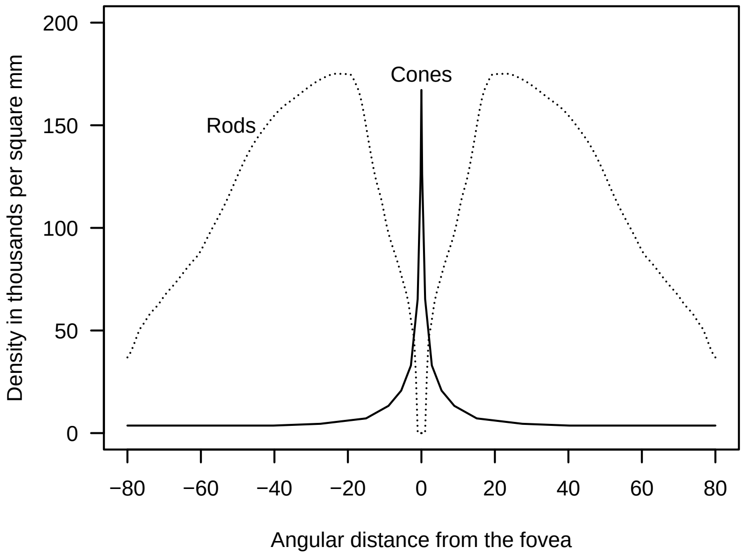

Figure 4.2: The distribution of rods and cones in the retina.

The distribution of rods and cones is not uniform across the retina. The cones are concentrated towards, and the rods away from the centre as shown in figure 4.2. In the middle of the retina is a small depression from 2.5 to 3 mm in diameter known as the yellow spot, or *macula*. At the centre of this is a tiny rod-free region about 0.3mm in diameter, called the *fovea centralis*. Within the fovea, the cone cells are very tightly packed together and the blood vessels and other cells are pulled aside to expose them directly to the light.

The concentration of cones in the fovea means that, in normal light, we have our best visual acuity in the centre of our visual field. The eye receives data from a visual field of about 200 degrees, however most of this field is perceived at low resolution because of the low-density of cones over most of the retina. To be seen at high resolution, the image of an object must fall on the fovea. This means that it can subtend an angle of no more than 15 degrees. This is just slightly larger than the image of the full moon. The very highest visual acuity occurs when the image falls on the fovea centralis which is perhaps on tenth of this size.

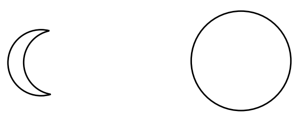

Figure 4.3: The image of the full-moon and the size of the fovea.

In dim light, such as that of a starlit night, the images we see come entirely from the rods in our eyes. Under these conditions, the fovea effectively acts a second blind spot. To see small objects at night, one must shift the vision slightly to one side, say 4 to 12 degrees, so that the light falls on some rods.

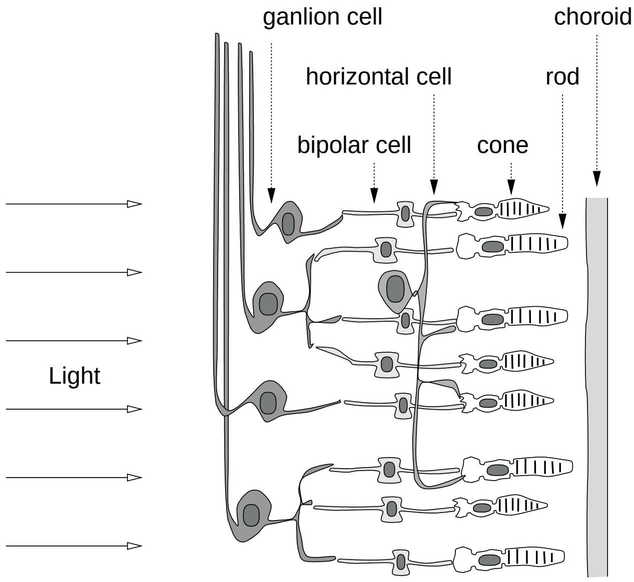

Figure 4.4: The arrangement of cells in the retina.

4.2.3 Retinal Circuitry

Although there are some 120 million rods and 6 million cone cells in the retina, there are less than a million optic nerve fibres which connect them to the brain. This means that there cannot be a single one-to-one connection between the photoreceptors and the nerve fibres. The number of receptors connecting to each fibre is location dependent. In the the outer part of the retina, as many as 600 rods are connected to each nerve fibre, while in the fovea there is an almost one-to-one connection between cones and fibres.

In addition to the rods and cones there are a number of other cell types whose function is to gather and process the information produced by the photoreceptors. The *ganglion cells* serve as terminators for the nerve fibres connecting to the brain. Between them and the photoreceptors are three other types of cells: *bipolar*, *amacrine* and *horizontal cells*. The bipolars receive and transmit signals from the receptors to one ganglion. Throughout most of the retina, bipolars gather signals from several receptors while in the fovea there is usually one for each cone. The horizontal cells connect adjacent receptors and amacrine cells link multiple ganglions.

Figure 4.4 shows the arrangement of these cells in the retina. Notice that the arrangement is counter intuitive, with light passing through the connecting "circuitry" before falling on the light sensitive receptors.

4.2.4 Lateral Inhibition

The complexity of the connections in figure 4.4 indicates the retina is capable of some quite complex signal processing operations. One form of processing which takes place in the retina is called *lateral inhibition*. When a local section of the retina is excited,

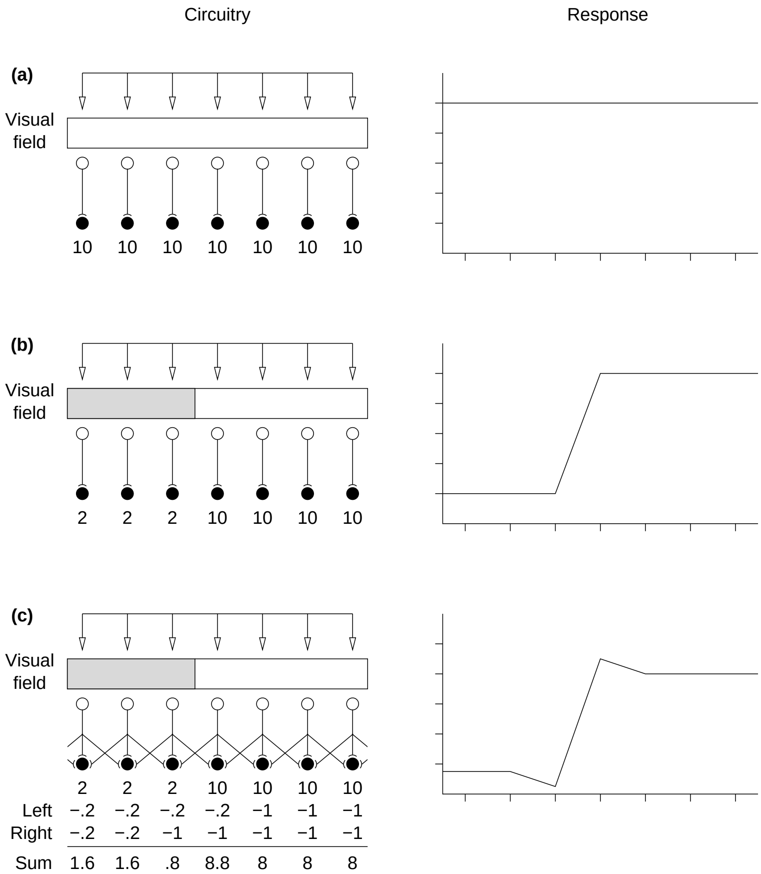

Figure 4.5: The effects of lateral inhibition.

the cells there do not just signal this to the visual processing centres of the brain. They also send signals to neighbouring cells whose effect is to diminish the effect of any excitation taking place there.

Figure 4.5 shows the effects of very simple model for lateral inhibition. Parts (a) and (b) of the figure show the sensory system operating without the effects of inhibition. In part (a) a uniform stimulation is applied to an array of sensors. The result is a constant level of output from the sensors. In part (b), the level of stimulation is decreased over the rightmost sensors. This has a direct on the output from the sensors.

Part (c) of the figure shows the effect when inhibition is introduced to the system. As well as outputting its signal, each sensor has an inhibitory effect on its two neighbours. The resulting sensor output is quite similar to (b), but at the boundary between the two levels of excitation, the difference in the output is accentuated. In image processing, treating signals in this way is known as *edge enhancement*.

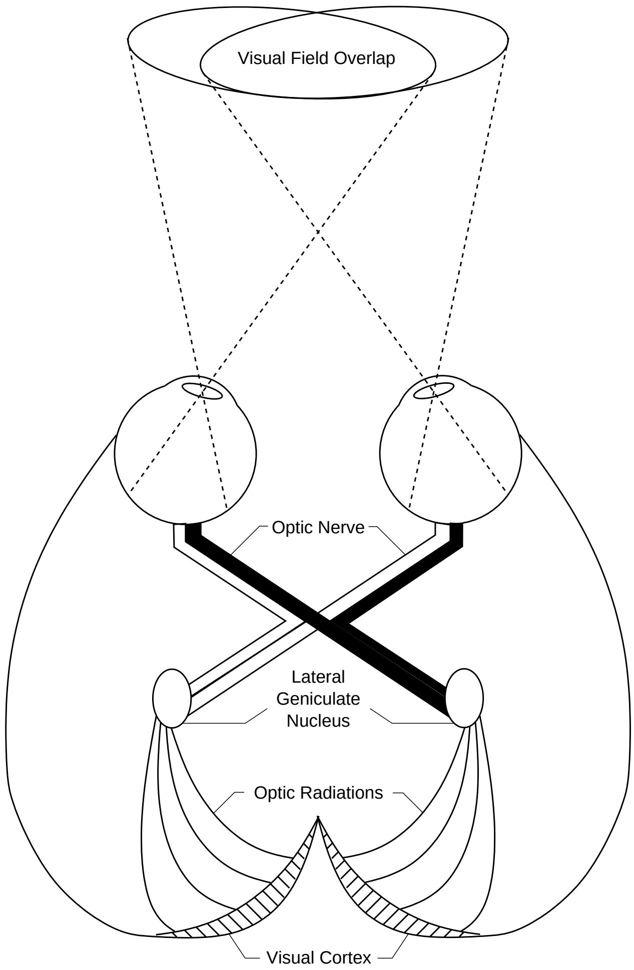

Figure 4.6: The human visual system.

4.3 Visual Processing in the Brain

Figure 4.6 shows the general structure of the visual system. The signals produced in the retina are propagated backward through the head along the optic fibre tract, which is linked via the lateral geniculate nucleus to the visual cortex. The signals from the right side of each eye are propagated to the right side of the brain and those from the left side of each eye to the left side of the brain. The images on the retina are reversed (see section 4.5.1) so that the right part of the brain processes the left part of the visual field and vice versa.

From the optic chiasma the optic fibres terminate in the lateral geniculate bodies. From these, optic radiations extend into the primary areas of the visual cortex. These in turn are connected to secondary visual association areas, and from there to the general interpretive area of the brain.

The six layers of the lateral geniculate nucleus are divided equally between information supplied from each retina and therefore provide the first stages of stereoscopic vision. The cell arrangements each lateral geniculate nucleus are similar to those of the ganglion cells of the retina, but far more are devoted to contrast and movement, and fewer to luminosity. Similarly he paired responses for red-green and yellow-blue are also found in the lateral geniculate nucleus, suggesting that its main function is to carry out further processing of incoming signals.

The cells of the visual cortex also have a spatial relationship with the ganglion receptive fields of the eyes. About 50% them are devoted to the macula, with the fields from the other retinal areas being represented in their correct spatial relationship.

The visual association areas of the brain appear to deal with more complex processing of the visual signal. In particular they appear to handle the processing of complex shapes and arrangements. Some evidence for this has been derived from people with brain damage in these areas, whose ability to identify or recognise shapes has been destroyed or severely impaired.

Although the lateral geniculate nucleus appears to concentrate on the contrast at edges and the discrimination of form, the primary visual cortex carries the process of form analysis even further. It appears that the cortex has columns of cells, extending through six layers, and particular directions. Some of these arrangements appear to be devoted to recognising lines or bars or particular lengths and at particular orientations.

The overall impression gained of how the brain processes visual information is as a series of steps, each taking the output of the previous steps and building up progressively more complex and abstract “impressions” of the input.

4.4 Eye Movements

When we look at an object, we only achieve high resolution for that portion of the object whose image falls directly on the fovea. Most objects we examine are much larger than this and to build a high resolution impression of them we must move our eyes so that all portions of their image fall, at least for a short time, on the fovea. These movements are carried in a series of jerky *saccades*, interspersed with stationary *fixations*. Both saccades and fixations last on the order of 100ths of a second.

Eye movements have been examined in some detail by a number of researchers. A particularly interesting account is given by the Russian researcher A. L. Yarbus [11]. Yarbus carried out experiments by attaching an apparatus (called a “cap”) directly to the eyeball of his subjects (the experiments are described in rather uncomfortable detail in Yarbus’ book). A small mirror was mounted on the cap reflected a light beam on to a sheet of photo-sensitive paper. When developed, this photographic paper yielded a record of the subjects eye movements.

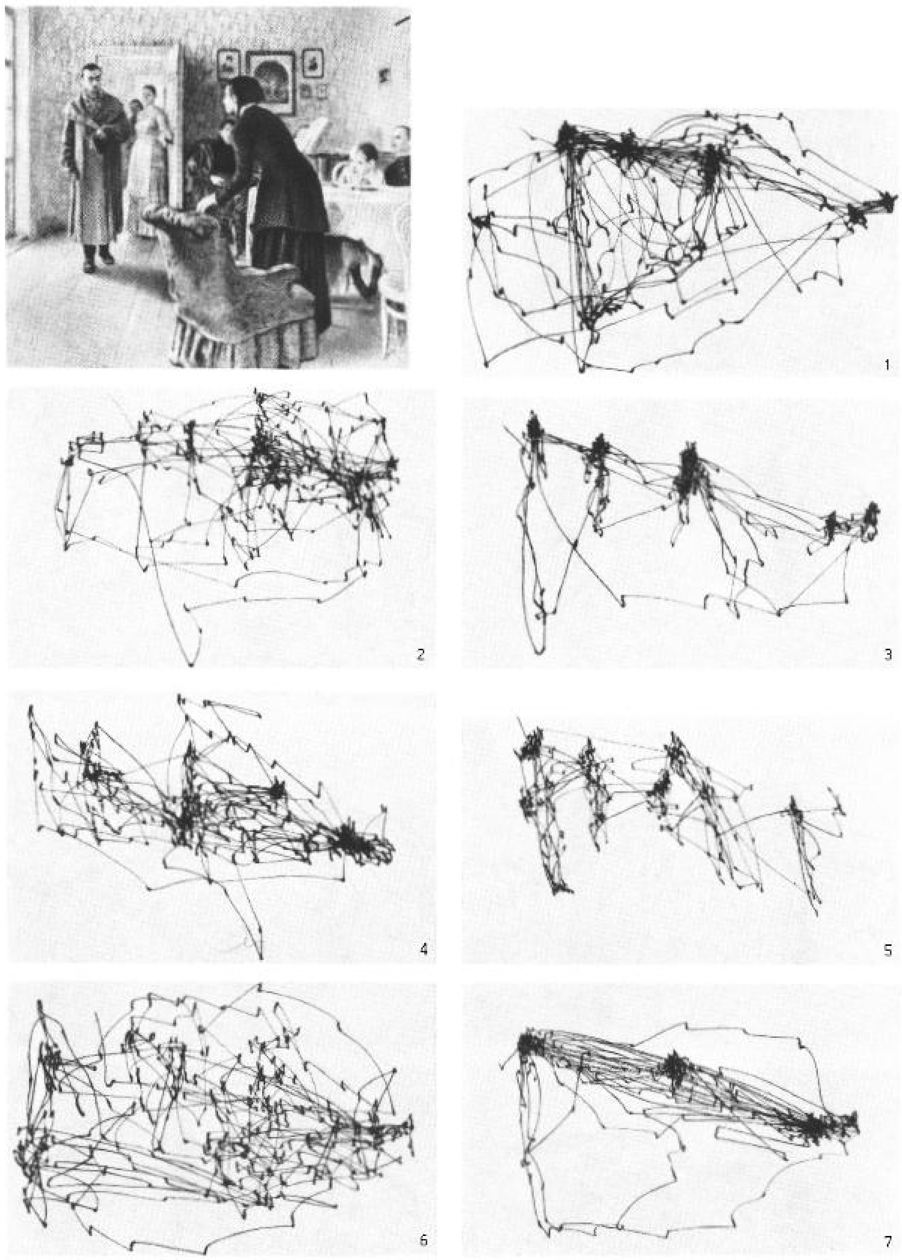

Figure 4.7: Eye movements directed by questioning.

In one part of his study, Yarbus showed a subject a reproduction of the picture *An Unexpected Visitor* by I. E. Repin. The subject was asked to:

1. Examine the picture with no suggested goal.
2. Estimate the material circumstances of the family in the picture.
3. Give the ages of the people.
4. Surmise what the family had been doing before the arrival of the unexpected visitor.
5. Remember the clothes worn by the people.
6. Remember the position of the people and objects in the room.
7. Estimate how long the "unexpected visitor" had been away from the family.

The pattern traced by the eyes and the points of fixation clearly depend on the goal given to the subject. Yarbus states (pp. 192–193):

> *Depending on the task in which a person is engaged, i.e., depending on the character of the information which he must obtain, the distribution of the points of fixation on an object will vary correspondingly, because different items of information are usually localised in different parts of an object. This figure shows that, depending on the task facing the subject, the eye movements varied. For example, in response to the instruction "estimate the material circumstances of the family shown in the picture," the observer paid particular attention to the women's clothing and the furniture (the armchair, stool, tablecloth and son on). In response to the instruction "give the ages of the people shown in the picture," all attention was concentrated on their faces. In response to the instruction "surmise what the family was doing before the arrival of the 'unexpected visitor,' " the observer directed his attention particularly to the objects arranged in the table, the girl's and the woman's hands, and to the music. After the instruction "remember the clothes worn by the people in the picture," their clothing was examined. The instruction "remember the position of the people and objects in the room," caused the observer to examine the whole room and all the objects. His attention was even drawn to the chair leg in the left part of the picture which he had hitherto not observed. Finally, the, instruction "estimate how long the 'unexpected visitor' had been away from the family," cause the observer to to make particularly intensive movements of the eyes between the faces of the children and the face of the person entering the room. In this case he was undoubtedly trying to the answer by studying the expressions on the faces and trying to determine whether the children recognised the visitor or not.*

Yarbus used this methodology to study a variety of questions about the way in which we extract information from pictures, including experiments see if visual illusions are caused by eye movements (he concludes that they are not).

Figure 4.8 shows the result of one perceptual experiment conducted by Yarbus. The subject was shown the picture in the top half of the figure and asked to trace the horizontal line from right-to-left and then the diagonal line from left to right, extending the second trace in the direction of the extension of the diagonal line. The lower part of the figure shows the resulting eye motion superimposed on the picture. It is interesting to note that the subject appears to overestimate the angle between the line to be followed and the horizontal.

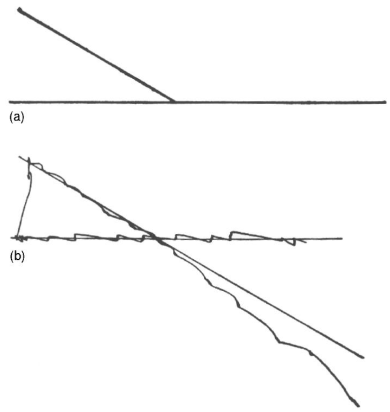

Figure 4.8: Eye motion when extrapolating an angle.

In another intriguing experiment, Yarbus examined what happens when eye motions are eliminated (by making the image being viewed move with the eye). In that case, after 1 to 3 seconds, the image was no longer visible to the subjects. When eye motion over the image was restored, the image reappeared.

Some attempt has been made to apply this kind of methodology to map reading, but no major study has been made of the way in which we extract information from graphs. Such a study could undoubtedly answer (and raise) many questions about this kind of activity.

4.5 Seeing In Three Dimensions

Of all our senses, it is sight which provides the most powerful perception of the three dimensional world around us and our visual abilities are finely tuned to handle the fact that we inhabit such a world. Despite the fact that we have strong spatial perception abilities, our basic visual hardware is two dimensional (we project a view of the world onto the retina). Understanding how we obtain a sense of depth from such a two dimensional system will be helpful when we consider visual perception in general.

To check that our perception is indeed three dimensional, have someone hold a pencil, point up, in front of you. With both eye open, you will find that you can easily reach out your hand and bring your finger down to touch the point of the pencil with uncanny accuracy. This task is considerably harder when repeated with one eye closed. This is because the differences in the direction of gaze of the individual eyes provide us with a form of triangulation which can be used to judge distance. (In this case it is unlikely that we know the precise distance to the point of interest, rather, we know when our finger is the same distance away as that point.)

There is no single way in which we gain our perception of depth. Some of it comes from the geometric properties of our visual system and some comes from experience

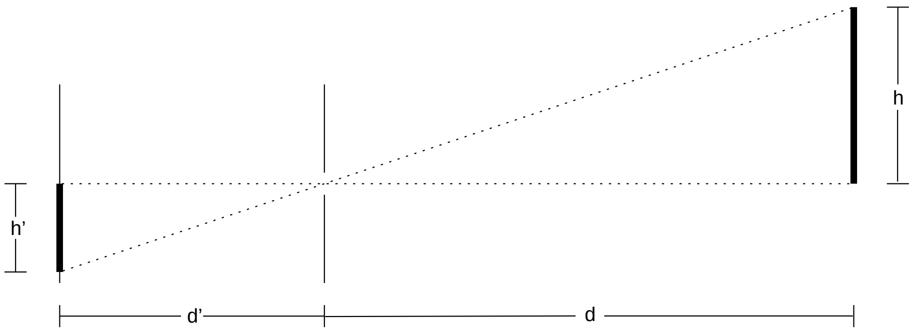

Figure 4.9: Geometry of a pinhole camera.

of how light illuminates objects in the world around us.

4.5.1 Perspective

The eye functions in the same way as a pinhole camera. Light from the scene being viewed passes through a small hole (the pupil) and is projected on a screen behind it (the retina). This arrangement has a major effect on the way we see.

Figure 4.9 shows the geometry of how an image is projected through a pinhole camera. Suppose that an object of height h is placed at distance d in front of the pinhole and projected through it onto a screen a distance d' behind the hole, resulting in an image on the screen that has height h'. Because the triangles are similar we know that

$$\frac{h'}{d'} = \frac{h}{d}$$

and hence

$$h' = \frac{d'}{d}h.$$

Assuming that the distance between the pinhole and the screen being projected on is constant, this says that apparent size of an object is inversely proportional to its distance from the observer. This dependence of the apparent size of objects on their distance away is know as *perspective*.

The fact that we see with perspective has a number of interesting consequences. If we look a pair of parallel lines, the apparent distance between the lines gets smaller as the lines recede into the distance. This means that such lines appear to converge at a *vanishing point* on the far horizon.

The discovery of the law of perspective was one of the major discoveries of renaissance painting. The feeling of depth that perspective yielded, produced a quantum leap in the level of realism of paintings, and artists used the device obsessively.

Figure 4.10 shows two images of a cube. The cube on the left is drawn using perspective, and that on the right is drawn without perspective — the corners are projected along parallel lines to the projection screen (an *orthographic* projection). The cube of the left looks more natural of the two, because of the feeling of depth which perspective gives. For the right-hand cube, there is even ambiguity about which faces of the cube are at the front and which are at the back.

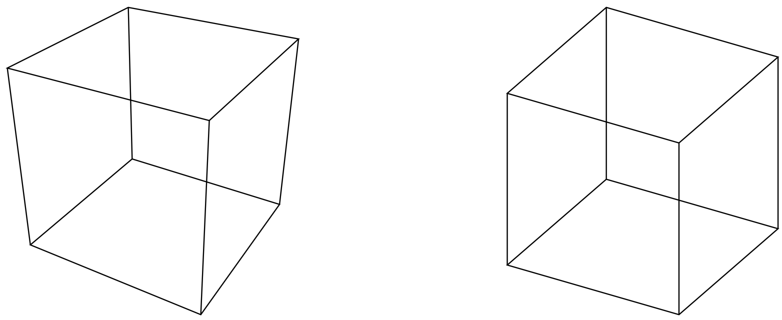

Figure 4.10: A cube drawn with and without perspective.

Perspective gives us very strong “depth cues” which enable us to to perceive depth in a two dimensional representation of a three dimensional scene. Unfortunately, our perception of truly two dimensional scenes can be incorrect when those scenes contain structures which can be interpreted as the result of perspective. For example, it is virtually impossible to interpret the left-hand cube in figure 4.10 as a set of straight lines on the paper. The temptation to “see depth” is just too strong.

The Ames Room

Even when viewing true three dimensional scenes it is possible to be mislead by perception. Figure 4.11 shows a picture taken of a very special room called the Ames room after its inventor, American ophthalmologist Adelbert Ames. Ames first constructed such a room in 1946, basing his design on a concept originally conceived by Hermann Helmholtz in the late 19th century.

A pair of twins is present in the room, but one appears much larger than the other. As the twins move about the room they appear to change size, shrinking as they move to the back left corner and growing as the move right and forward. The room has other mysterious properties too. It it possible to make balls appear to run up ramps in defiance of gravity.

The Ames room is an example of a very carefully crafted illusion. The room is not rectangular as it appears, but is trapezoidal, with the back left corner being much further away from the viewer than the back right corner (as shown in figure 4.12). It is also much higher at the back left corner than elsewhere. The shape of the room is very cleverly disguised by a variety of cues which use perspective to suggest a rectangular shape.

There is only one viewing position which makes the illusion work. The viewer is forced to use the position because the only way to look into the room is through a small peephole placed at exactly the right position. Even when this position is used, the illusion is not completely convincing because there are other cues which suggest conflicting depths.

4.5.2 Stereoscopic Vision

Perspective is probably the strongest depth cue we have for objects which are a relatively large distance away from us. For objects which are close, there is another way

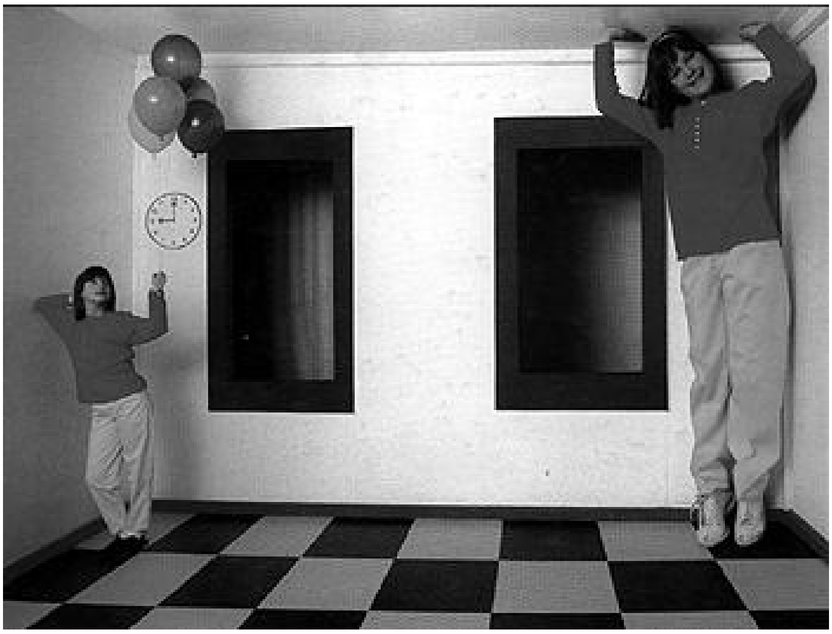

Figure 4.11: The Ames room.

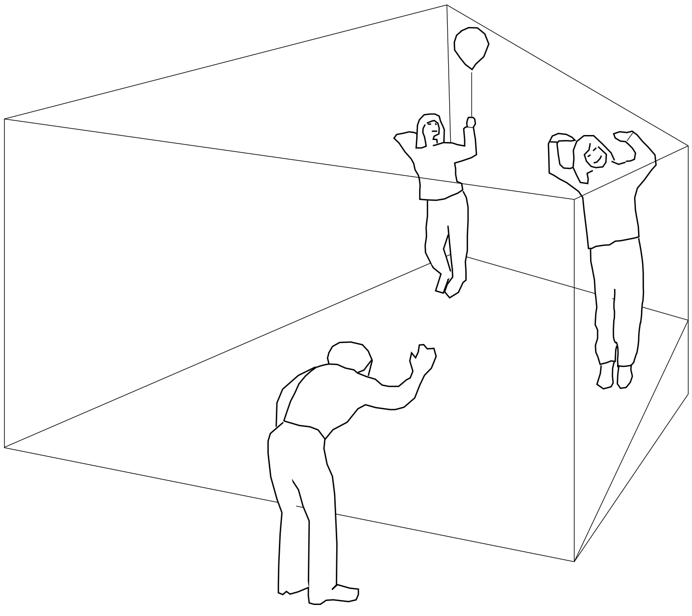

Figure 4.12: A schematic view of the Ames room.

of judging distance which is useful; *stereoscopic vision*.

We we look at objects with both eyes open we actually receive two slightly different images of them because our eyes occupy sightly different positions relative to those objects. Our brains make use of the differences between the two images to infer depth information. The process is completely unconscious and we are usually unaware that we are performing it. An exception to this happens when an individual loses sight in one eye. They find that they sometimes “miss” when they attempt to place something on a shelf or table, at least until they develop compensating strategies.

The experience delivered by stereoscopic vision is far richer that obtained by looking at a flat image. Because of this there are a number of technologies which try to deliver images stereoscopically. These range from the old fashioned stereopticon (a kind of 3d slide viewer) through red-green filter glasses for watching movies and television, to immersive virtual reality environments. These all work by presenting slightly different flat images to each of the two eyes, thus fooling the brain into interpreting the images as having depth.

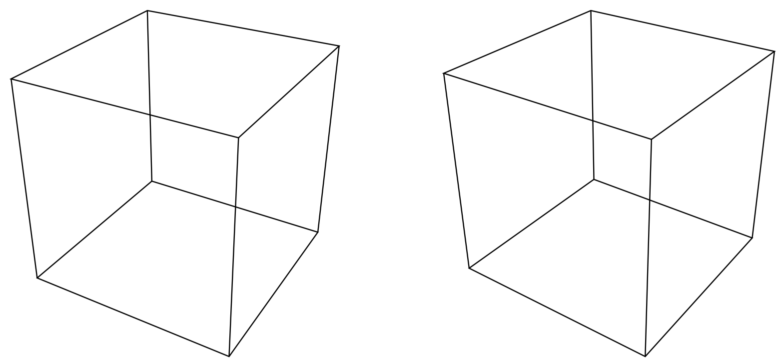

Figure 4.13: Stereoscopic views of a cube.

Figure 4.13 shows an example of the two different views our eyes would have of cube, if it was placed at about an arm’s length from our eyes. The two views are quite similar, but are different. If we present these two views separately to our eyes, we see a view of the cube which is truly three dimensional.

There are a number of ways to present separate views to the eyes. The simplest is to use a stereopticon. This is a device which has two slightly magnifying lenses mounted to look through at a matched pair of images such as those in 4.13. The device has a separating card running from between the eyes to between the images to ensure that the views for each eye are kept separate.

With a little practice it possible to merge the images without the aid of a stereopticon. The trick is to look through the page to bring the images into alignment before attempting to bring them into focus.

4.5.3 Occlusion

We learn at a very early age that it is possible for objects which are close to hide or *occlude* more distant objects. It is this fact that determines how painters produce their pictures. Start with the most distant objects and progressively work forward to the nearest objects.

4.5.4 Lighting Effects

We gather a lot of information about the structure of the world by studying the way that light falls on objects. In the figures below, we interpret the shadows in a way which makes the circles either protrude from the page or recede into it. Usually we interpret scenes as being lit from above and this determines which circles come protrude and which recede. Turning the paper upside down reverses the impression.

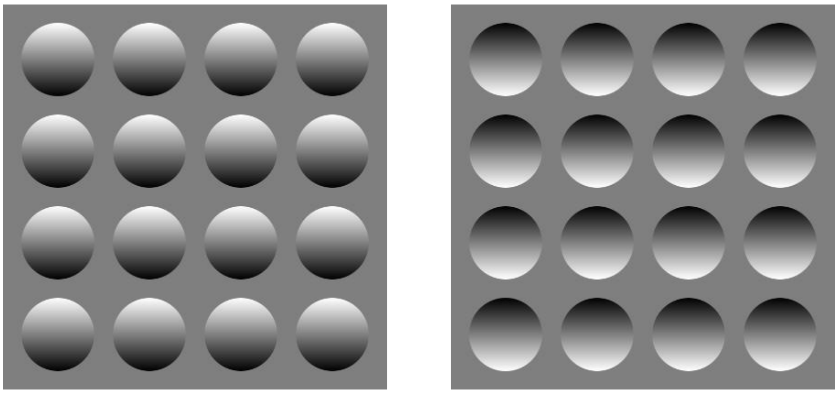

Figure 4.14: Structure revealed by light.

The use of shading effects can help convey a sense of depth to a picture. This was recognised by painters and used to help provide a sense of realism in their paintings. The work of Rembrant shows some wonderful examples of this kind of use of light and shade.

The effect of light falling on objects is also been used in a number of visualisation applications. One of the most common is the use of shading in topographic maps, where it is known as *hill-shading*.

4.5.5 Haze and Fog

If you stand somewhere where you can see far off into the distance, you will notice that the far-off hills are lighter or more washed-out. This effect is caused by haze in the air scattering the light from the more distant objects. This phenomenon has been used to provide depth cues, particularly in high-end graphics system. the graphics are computed so that more distant parts of the image are lighter and less distinct.

This kind of depth cue was fashionable in high-end CAD systems for a while, and it is widely used by landscape painters to produce the illusion of depth in their pictures.