An Introduction to 3D Computer Graphics, Stereoscopic Image, and Animation in OpenGL and C/C++

Fore June

Chapter 8 Depth Perception

8.1 Stereoscopic Depth Perception

When we observe the three dimensional world with our eyes, the three dimensional environment is reduced to two-dimensional images on the two retinae of our eyes as shown in the figure below.

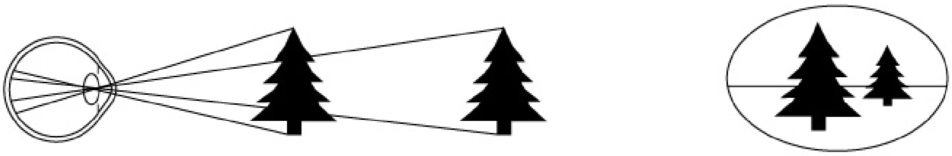

Figure 8-1 3D to 2D Projection on Retina
(Image from *http://www.mind.ilstu.edu/curriculum/vision_science_intro/vision_science_intro.php*)

We usually call the "lost" dimension *depth*, which is crucial for almost all of our interaction with our environment. Our brain tries to estimate the depth from the retinal images as good as it can with the help of different cues and mechanisms. *Depth perception* is the visual ability to perceive the world in three dimensions (3D) and the distance of an object. *Depth sensation* is the ability to move accurately, or to respond safely, based on the distances of objects in an environment. In nature, depth sensation is crucial for the survival of many animal species.

The main reason for humans to have depth perception is that we have two eyes located in our forehead and the two eyes are separated by a distance of about 2.5 inches. Because the left eye and right eye are at different positions, the 2D images of the 3D environment projected onto their retinae are not the same. The slight difference in the two projected images can be utilized to extract depth information. We usually refer to the discrepancies perceived by the two eyes as **binocular disparities**. Our brain converts binocular disparity to depth information based on our experience.

It turns out that eagles have a similar distance between the eyes as humans do as shown in the figure below. Therefore, eagles also can sense depth well. Actually, eagles have better vision than humans at the day time because their eyes have better cone cells.

Figure 8-2 Eagle Eyes and Human Eyes

Some readers may wonder whether some small animals with two eyes close to each other may have difficulties to extract depth information from the binocular disparities as the two projected images will be very close to each other. Indeed, animals such as insects and small birds whose eye separations are very small cannot resolve depth from binocular disparities; they have to rely on other methods such as object motions to infer the distance of an object. It turns out that even if we see with one eye, we still can extract a lot of depth information based on our perspective experience of the world. Actually, the problem of finding depth from 2D retinal images is an under-determined problem. However, in most situations we can perceive depths of objects in 2D images with ease.

What information contained in a 2D image makes us perceive depth? This approach of correlating information in the retinal image with depth in the scene is called *cue approach*. This approach tries to relate the elements of information contained in a 2D scene to the depth of the objects in the scene. According to cue theories, we make a connection between these cues and the actual depth by our accumulated knowledge from prior experience. The association becomes automatic after repeated experience of observing this 3D world. These include some heuristics, which may confuse us in pathological cases. We can roughly divide depth cues in a few different categories:

1. **Oculomotor**: Cues based on our ability to sense the tension in the eye muscles.
2. **Monocular**: Cues that work with one eye.
3. **Motion-based**: Cues that depend on motion of objects.
4. **Binocular**: Cues that depend on two frontal eyes.

8.2 Oculomotor Cues

Convergence and accommodation of our eyes can induce 3D information. *Accommodation* is the process by which our eye changes shape to maintain a clear image (focus) on an object as its distance changes. When looking at nearby objects, our eyes move inwards. This is called *convergence*. When looking at far away objects, our eyes move outward. Our eyes are relaxed and are non-convergent as shown in the right image of the figure below.

Figure 8-3 Converged Eyes and Non-converged Eyes

When the eyes move inward, there is a tightening of the muscles that hold the lens to get the eye to focus close by (left of Figure 8-3). When the eye moves outward, muscles relax (right of Figure 8-3). These two situations give us different sensations of our eye muscles which give us an estimate about the depth of the object. We can experience this by moving an object at our arms length closer and closer to the eye to feel the tightening of the muscles and then move it back again to feel the muscles relax. Actually, we can relax our eye muscles and thus improving our eyesight by alternating our staring at nearby and far away objects over a short period of time. So playing table tennis or badminton are good for our eyes.

There is a direct relationship between the effort of accommodation and the effort of convergence. We measure the effort of accommodation in *diopters*, and the effort of convergence in *degrees*. A diopter, or dioptre, is a unit of measurement of the optical power of a lens or a curved mirror. It is equal to the reciprocal of the focal length measured in metres. Therefore, its unit is the reciprocal of length. For example, a 3-diopter lens brings parallel rays of light to focus at $\frac{1}{3}$ meter from the lens.

There are three main types of eye refraction: emmetropia (normal-sightedness), myopia (nearsightedness), and hyperopia (farsightedness). Emmetropia is the normal condition of perfect vision, where parallel light rays are focused on the retina without the need for accommodation. When an emmetropic person is looking at a distant object, accommodation and convergence are at in a resting state; the axes of the eyes are parallel, and there is no effort of convergence (the angle of intersection is zero degrees). When the emmetropic (normal sighted) person is looking at a nearby object, some effort of accommodation and convergence is necessary. When an emmetropic person is looking at the distance of one meter, the effort of accommodation is 1.00 diopter, and if the eye separation is $65mm$, the angle of convergence θ is:

$$\theta = tan^{-1}\frac{65}{1000} = 3.7^o$$

Similarly, at a distance of 0.5 meters, if the effort of accommodation is 2.00 diopters, the angle of convergence is about 7.4 degrees. At a distance of 0.33 meters, if the effort of accommodation is 3.00 diopters, the convergence is about 11 degrees. Figure 8-4 shows the convergence of a nearby object and the convergence of a far object.

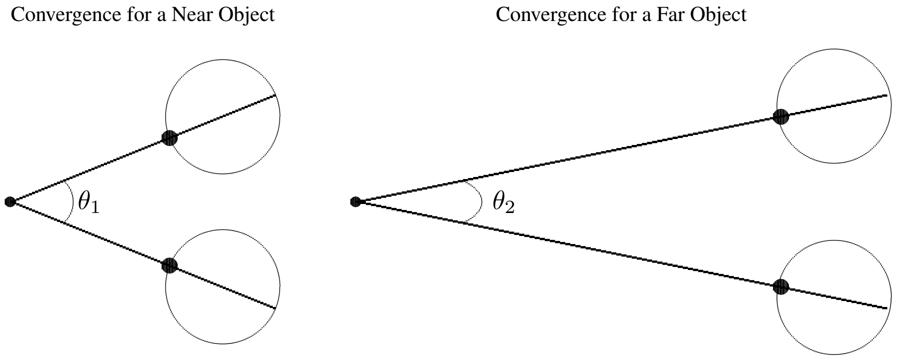

Figure 8-4 Angles of Convergence ($\theta_1 > \theta_2$)

8.3 Monoscopic Depth Cues

We can extract a lot of 3D information from a single 2D view. As we have discussed above, the 3D world is projected into a 2D image on the retina of an eye. People with one blind eye can only detect 2D images. However, you might know of some people with one blind eye who still can sense this three dimensional world. They might behave like a normal person and you might not be even aware of their handicap. They can move to a chair and sit properly on it; they do not bump into a wall and drop glasses over the table edges. They achieve these by extracting 3D information from a single 2D image they receive. The depth cues that are obtained from the 2D image of only one eye are called *monocular cues* or *pictorial cues*, which usually do not provide absolute information about depth, but relative depth with respect to other objects in the environment. Monocular cues include the following different types.

8.3.1 Occlusion

When an object is hidden fully or partially, this hidden (occluded) object is considered to be farther away, behind the object (occluding object) which is covering it. For example, if a house hides half a man, we know that the house is in front of him. *Occlusion* is also referred to as *interposition*. Figure 8-5 shows that the square is behind the circle as the circle hides part of the square.

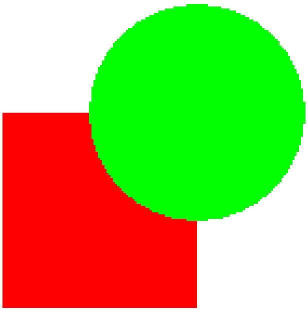

Figure 8-5 Occlusion

Figure 8-6 also shows how depth cues can be obtained from occlusion. Sometimes, the depth information they present may be ambiguous as shown in Figure 8-6(b). Actually, many common optical illusions are based on these ambiguities. Despite the potential for ambiguity, combining many depth cues produces a powerful sense of three dimensionality. Figure 8-6(c) shows that the shadow cues resolve contradicting depth cues deduced from occlusion and relative object heights.

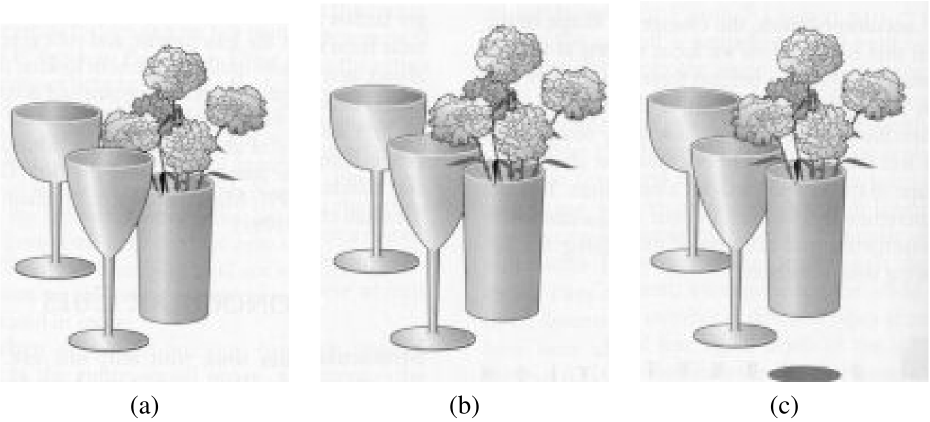

Figure 8-6 (a) Occlusion Cues
(b) Contradicting Occlusion and Relative Height Cues
(c) Shadows Resolving Contradiction

8.3.2 Relative Height

Relative height, also called *height in field*, is another cue to infer depth. We judge an object's distance by considering its height in our visual field relative to other objects. The usual assumption is that where the base of a shape is higher than that of a similar shape, then the one with the higher base is farther away. This cue relates objects to the horizon. As objects move away from us, objects get closer to the horizon. Therefore, we perceive objects nearer to the horizon as more distant than objects that are farther away from the horizon. This effect is shown in Figure 8-1 above and Figure 8-7 below. This means that below the horizon, objects higher in the visual field appear farther away than those that are lower. However, above the horizon, the effect is reversed; objects lower in the visual field appear farther away than those that are higher. These effects are shown in Figure 8-8. The clouds higher up in the image appear to be closer as they are above the horizon.

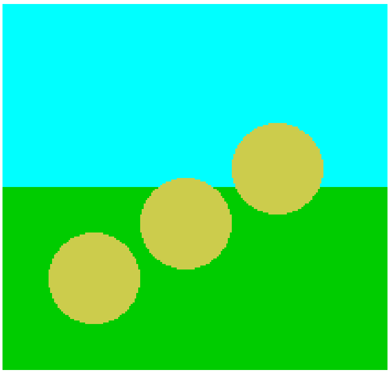

Figure 8-7 Distance to Horizon Induces Depth

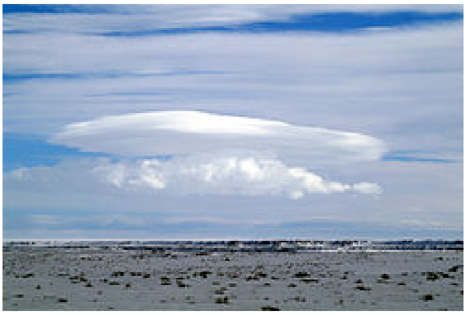

Figure 8-8 Clouds Higher Up in Sky Appear Closer

8.3.3 Aerial Perspective, Atmosphere Blur, Saturation, and Color Shift

If the air is smoggy, particles suspended in the air make distant objects look blur like the images shown in Figure 8-9 below. Moreover, relative colours of objects give us some cues to their distances. Diffusion of water particles in the air will generate a bluish color shift and pollution-based diffusion will generate grey to brown color shifts.

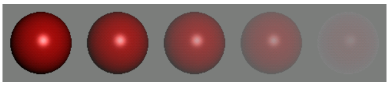

Figure 8-9 Smoggy Atmosphere

Due to the scattering of blue light in the atmosphere, distant objects such as distant mountains appear more blue. Contrast of objects also provide clues to their distance. When the scattering of light blurs the outlines of objects, the object is perceived as distant. Mountains and objects are perceived to be closer when they appear clear. These effects are shown in Figure 8-10. Figure 8-11 shows a Chinese landscape painting using *atmospheric perspective* to show recession in space. Figure 8-12 shows aerial perspective as viewed towards a sunset; scattering has resulted in a shift of colour towards red.

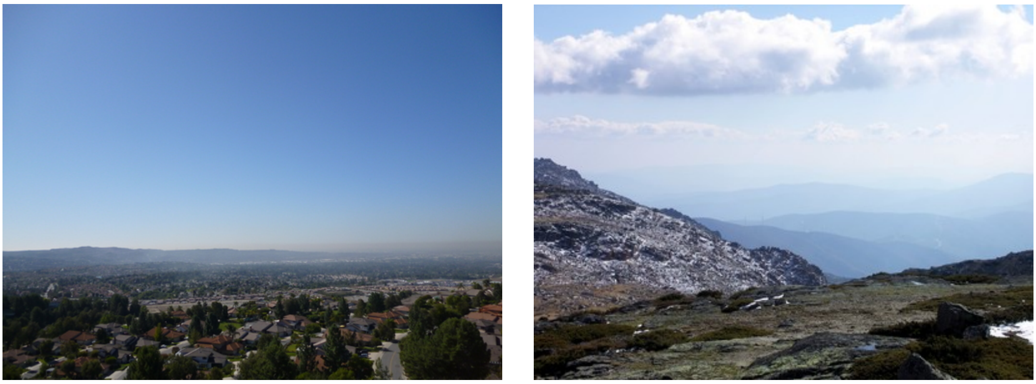

Figure 8-10 Aerial Perspective

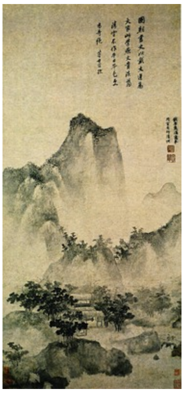

Figure 8-11 Atmospheric Perspective in Painting

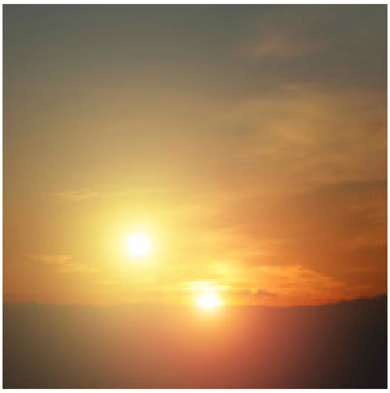

Figure 8-12 Color Shifting

8.3.4 Shadows and Highlights

We can extract depth information of an object from highlights and shadows. Because our visual system assumes light comes from above, a totally different perception is obtained if an image is viewed upside down. Figure 8-13 shows two rows of shadowed circular shapes which appear to be bumps on the left three columns; the illustration of the right three columns is obtained by viewing the left three columns upside down. The bumps become holes!

Figure 8-6 above shows an example where the occlusion and the relative height cues contradict each other. Shadows play an important role in this case to resolve the contradiction. The presence of the shadow removes the contradiction by telling us that the vase is in front of the glass but at a higher position.

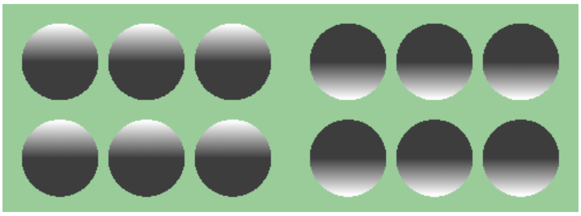

Figure 8-13 Right-Half is Upside-Down View of Left-Half

8.3.5 Linear Perspective

Lines in the 3D environment that are perpendicular to the retinal image plane and are parallel will not be projected into parallel lines in the 2D image. Rather, the lines tend to merge with increasing distance and seem to converge to a point at infinity, which is referred to as a **vanishing point**. Consequently, when common objects with parallel-line features such as roads, railway lines, and electric wires subtend a smaller and smaller angle in a 2D image, we interpret them as being farther away. This linear perspective helps us obtain depth cues. Figure 8-14 shows an example of this effect; the road seems to converge to a single point at infinity.

The simplest version of such illusions deriving from linear perspective is called the *Ponzo illusion*, which is an optical illusion first demonstrated by the Italian psychologist Mario Ponzo in 1913. Ponzo suggested that the human mind judges an object's size based on its background. He showed this by drawing two identical lines across a pair of converging lines, similar to those shown in Figure 8-15; in the figure, even though both horizontal lines are of the same size, we perceive the upper line as longer. This is because we interpret the converging sides according to linear perspective as parallel lines receding into the distance. In this context, we interpret the upper line as if it were farther away. So it appears longer to us as a farther object would have to be longer than a nearer one for both to produce retinal images of the same size.

Some researchers believe that the Moon illusion is an example of the Ponzo illusion. The Moon illusion is a phenomenon where the Moon appears larger near the horizon than it does when it is higher up in the sky; in this situation, the trees and houses play the role of Ponzo's converging lines, and foreground objects which trick our brain into thinking the moon is bigger than it really is. There are many other explanations of the Moon illusion. Some believe its due to the variations of the angular size when the moon is at different positions and some believe its due to the atmospheric effects. Ponzo illusion may just account for part of the magnitude in Moon illusion.

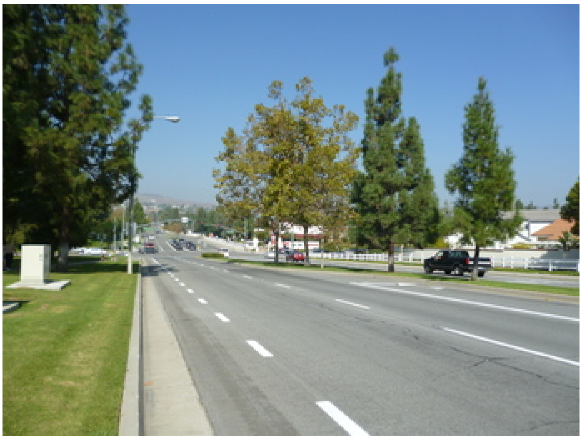

Figure 8-14 Linear Perspective

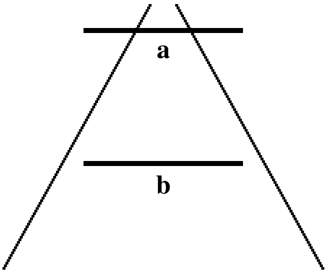

Figure 8-15 Ponzo Illusion

8.3.6 Relative Size and Familiar Size

Retinal image size allows us to judge distance based on our past and present experience and familiarity with similar objects. For example, if a building and a lady look the same size, we can assume that the building is farther away because from our real-life experience, a building is much larger than a lady. If two objects are equal in size, one that is farther away will have a smaller field of view than the closer one. We can combine this information with our previous knowledge of the object's size to determine the absolute depth of the object. For example, we are generally familiar with the size of an average automobile. We combine this prior knowledge with information about the angle it subtends on the retina to determine the distance of an automobile in a scene. This effect is shown in Figure 8-16; as a car drives away, the retinal image of it becomes smaller and smaller. We interpret a retinal image of a small car as a distant car.

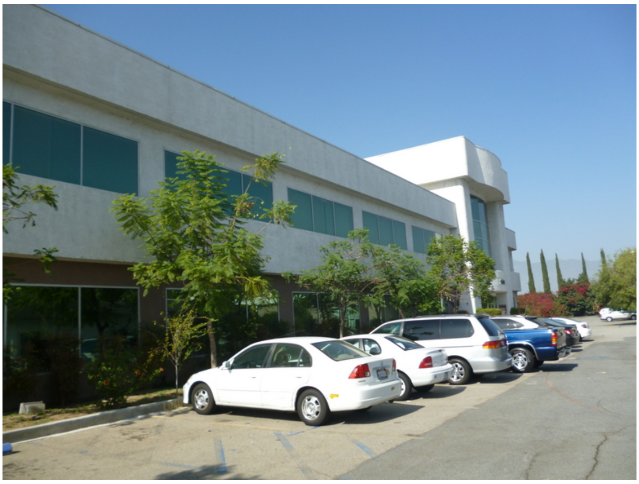

Figure 8-16 Familiar and Relative Size Cues

8.3.7 Texture Gradient

If we view a texture surface from a slanting direction, rather than directly from above, we will see a *texture gradient* meaning that the texture pattern does not appear uniform but appears smaller at distant positions. Though texture gradient relates in a sense to relative size, it is a depth cue in its own right. Many surfaces in the real world, such as walls and roads and a field of flowers in bloom, have a texture. The texture becomes denser, smoother and less detailed as the surface gets farther away from us and recedes into the background. This information helps us to judge depth.

Figure 8-17 shows an example of texture gradient. It is a photo image of a dried river bank. The diamond-shaped dried mud blocks are the key patterns. The mud blocks get progressively smaller as the river bank recedes in depth; eventually the blocks are not clearly distinguishable from each other. In the distant surface, only the general roughness of the river bank is noticeable and then this roughness becomes less noticeable at farther distance.

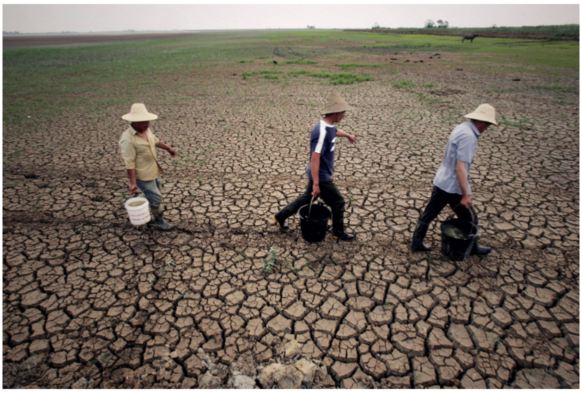

Figure 8-17 Texture Gradient

8.3.8 Familiar Shapes

We store the information of the shape of an object that we have encountered in our brain. When we see the object again, we will retrieve the shape information of the object from our memory. The object shape can give depth cues of the object. Figure 8-18 shows two objects of similar texture. However, we will perceive the left object as a convex soccer ball, until it skews and we identify it a a ball-textured concave plate shown on the right.

Shapes alone may not give strong depth perception. However, when we combine the knowledge of the shape of an object with other depth sense, we obtain strong depth cues. Studies have suggested that under many circumstances, the brain combines cues using weighted linear combination, where individual cues are first processed largely independent of each other to yield an estimate of an environmental property. The brain combines estimates from different depth cues by weighing each source of information proportional to its statistical reliability.

8.4 Motion-based Depth Cues

By analyzing the movements of objects, our brain can reveal the objects' speeds and directions, as well as their placement in the 3D environment. We can estimate how far away a moving object is by watching how fast it seems to move. For example, an airplane in sky looks stationary but appears to move very fast on the runway. If information about the direction and velocity of movement is known, motion *parallax*, which is the relative position of an object's retinal image at various times can provide absolute depth information. Some small animals such as birds rely on motion parallax to determine depth.

On the other hand, our head movements change our viewpoints of the environment frequently. These point-of-view (PoV) movements also cause motion parallax by which images of nearby objects cross the retinal image plane faster than those of distant objects. We observe this phenomenon when we sit inside a train and look outside through a window like the situation shown in Figure 8-19; we see that nearby objects speedily pass by in a blur, but distant objects move slightly. We can use the speed of these passing objects to estimate the distance of the objects.

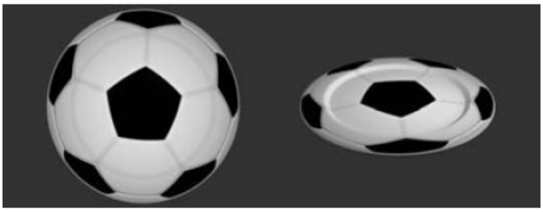

Figure 8-18 Depth Cues from Shapes

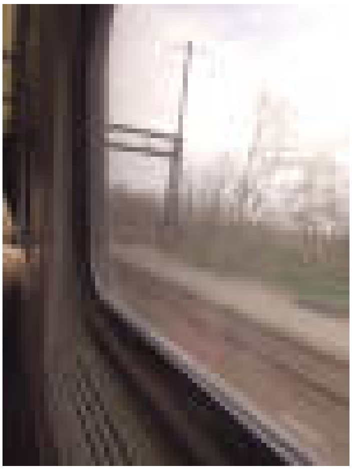

Figure 8-19 Motion Parallax
(From *http://www.morguefile.com/archive/display/1696*)

The sequence of images in Figure 8-20 below shows another example of motion parallax induced by PoV movement. As the viewpoint moves side to side, the displacements of nearby objects in Figure 8-20 are much larger than those of distant objects. When the sequence of images are shown as consecutive frames in a movie or a video game, the distant objects will appear to move more slowly than the objects close to the camera. Moreover, our eye movement also leads to deletion and accretion of objects as shown in the figure. As we move, parts of objects get revealed or occluded. We obtain depth information of the objects from the rate of these deletion and accretion.

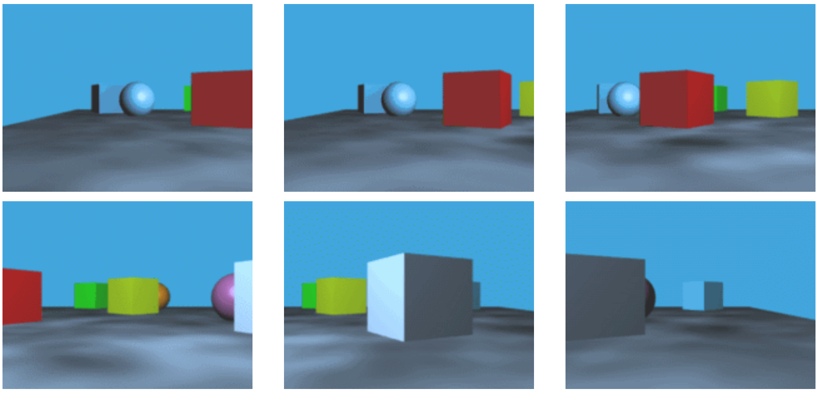

Figure 8-20 Motion Parallax Example
(Images from *http://en.wikipedia.org/wiki/Parallax*)

8.5 Binocular Cues

Animals that have frontal eyes can use information derived from the different projection of objects into each retina to deduce depth. The two eyes of a human are about 7 cm apart and hence they view the same scene with slightly different angles. Consequently, the projected 2D retinal images on the two eyes are different from each other. Though there is a significant overlap of view from the two eyes, their views are not identical as their locations are different. This difference in view is called *binocular disparity* and the brain converts it to depth information. The information provided by the disparity is called *stereopsis*.

The distance to an object can be derived with a high degree of accuracy using a process called *triangulation*, in which we determine the location of a point by measuring angles to it from known points at either end of a fixed baseline. In other words, given a baseline and two angles measured from the baseline, we can fix the whole triangle. Figure 18-21 shows how this is done. Suppose e is the known distance between the two points A and B. (For example, e can be the eye separation.) Then the distance d of the third point C from the baseline AB can be determined by trigonometric identities.

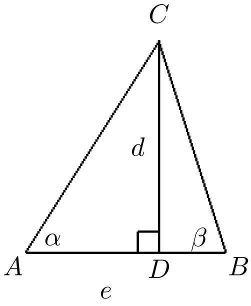

Figure 18-21 Triangulation

That is,

$$\begin{aligned} e &= AD + DB \\ &= \frac{d}{\tan\alpha} + \frac{d}{\tan\beta} \\ &= d[\frac{\cos\alpha}{\sin\alpha} + \frac{\cos\beta}{\sin\beta}] \\ &= d[\frac{\cos\alpha\sin\beta + \cos\beta\sin\alpha}{\sin\alpha\sin\beta}] \\ &= d[\frac{sin(\alpha+\beta)}{\sin\alpha\sin\beta}] \end{aligned}$$

Therefore, the distance d to the baseline is given by

$$\boxed{d = \frac{e \times \sin\alpha\sin\beta}{sin(\alpha+\beta)}} \tag{8.1}$$

We may be able to triangulate the distance to an object with a high degree of accuracy. Equation (8.1) shows that one can derive the distance of an object from the eye separation and the viewing angles of the eyes. However, it does not show the relation between the object distance and the disparity of the retina images. In reality, if an object is far away, the disparity of the images projected on both retinas is small. If the object is near, the disparity is large. It is *stereopsis* that tricks us into thinking of perceiving depth when viewing a 3-D movie or a stereoscopic photo.

One popular approach to generate stereo displays in cinema is to use different polarized light to project the images from two different viewpoints on the same screen. The viewers are provided with polarized glasses. The left glass allows one direction of polarization and the right glass allows the other. Though a viewer is watching the same screen, the two eyes receive two different images. The right image is blocked by the left glass and the left image is blocked by the right glass. This creates a compelling sense of depth. A simpler approach is to use colored glasses; the left glass allows one color such as red to pass through and the right glass allows another color such as blue to go through; the viewers are presented with two images of the same scene from two viewpoints, one created using red color and the other using blue color leading to a depth sensation. Similar technologies are used in head-mounted displays where a head gear with two micro-screens, one in front of each eye, is mounted on the head of the viewer. A sense of depth is created by projecting two images of the same scene generated from two different viewpoints on two different screens.

8.5.1 Horopter

Horopter, derived from the Greek words *horos* (boundary) and *opter* (observer), is a 3D curve used for modeling the human visual system. It consists of the points for which the light falls on corresponding areas in the two retinas. This is the same as the set of points of intersection of a *cylinder* and a *hyperbolic paraboloid.* A horopter can be regarded as a one-turn helix curve around a cylinder like the one shown in Figure 8-22. In the special case that he horopter is in the horizontal plane through the eyes that contain the centers of the retinas, it is referred to as the *Vieth-Muller circle* . In this case, the horopter takes the form of a circle, plus the line along the cylinder, and it is referred to as the *vertical horopter*. The center region of the retina, where the light falls on when we look straight to an object, is called the *fovea*. The eye structure is shown in Figure 8-23; light enters the pupil and is focused and inverted by the cornea and lens; it is then projected onto the retina at the back of the eye where seven layers of alternating cells process and convert the light signal into a neural signal.

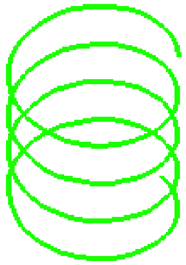

Figure 8-22 Helix Curve

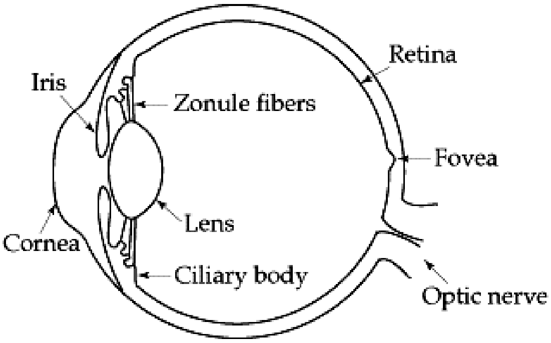

Figure 8-23 Eye and Retina

The Vieth-Muller circle is the theoretical horopter, which can be computed based on the geometry of the eyes and the viewing distance. The more general horopter of an individual is found empirically, where the horopter is the locus of points in space that appear binocularly fused to the observer; all the points lying on the horopter are imaged at corresponding points in the two eyes. *Corresponding retinal points* are the locations in each retina connected to the same place in the visual cortex; they give rise to the same visual direction. We can determine these points by locating the matching points on the retina when one retina could be slid on top of another. Anywhere else in space appears as a double image to the observer. The actual horopter measured in a laboratory is referred to as *empirical horopter*. These are shown in Figure 8-24. In the figure, the fixation point is the point in the visual field that is fixated by the two eyes in normal vision and for each eye is the point that directly stimulates the fovea of the retina. The nodal point is the point in the eye where light entering or leaving the eye and is undeviated. This allows similar triangles to be used to determine the retinal image size of an object in space. The horopter varies across individuals and with fixation distance and gaze angle.

The images of points closer than the horopter do not lie on corresponding points across the retina. Moving an object slightly off the horopter creates a small amount of *retinal disparity* that gives rise to stereoscopic depth perception as shown in Figure 8-25.

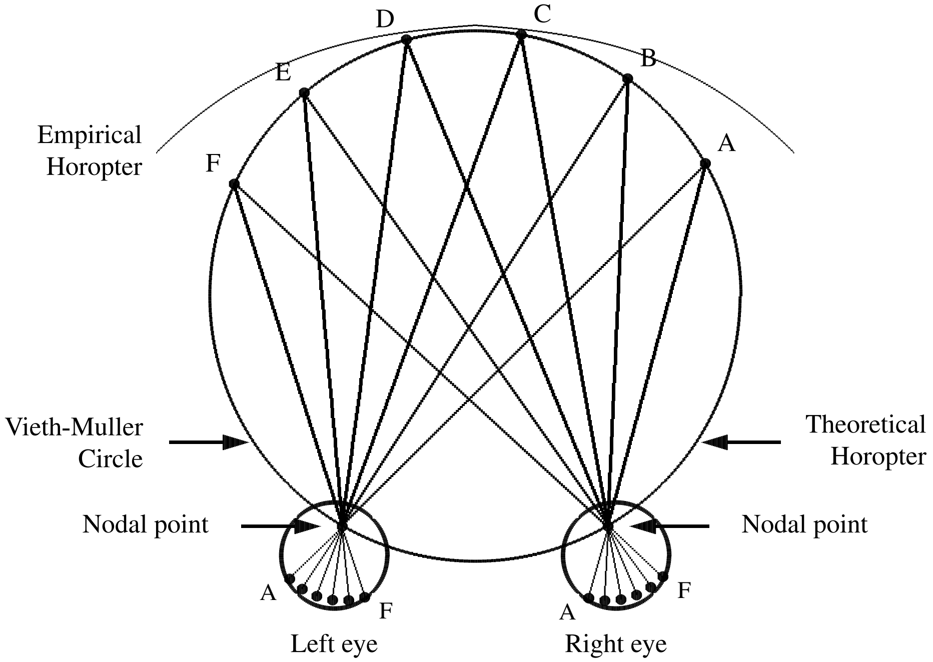

Figure 8-24 Horopter

There are physiological evidence that we are sensitive to retinal disparity. Researchers have identified cells in the visual cortex that respond selectively to different ranges of disparities. Though humans can convert retinal disparities to depth information with utter ease, it is not easy to devise computer vision algorithms to do so.

Note that only animals with frontal eyes such as cats, dogs, and humans make use of disparity to perceive depth. Some animals such as rabbits, buffaloes and birds which have lateral eyes located on both sides of the head cannot use disparity to sense depth. However, those animals have a wider view of the world than animals with frontal eyes. In such animals, the eyes often move independently to increase the field of view. Even without moving their eyes, some birds have a 360-degree field of view. They have much better ways of using motion cues to sense depth. In the process of evolution, predators like eagles grow frontal eyes as depth perception is very important for them to hunt successfully. On the other hand, animals lower in the food chain, herbivores such as rabbits and deers usually grow lateral eye so that they can have a larger field of view to see a larger part of the environment to detect predators more effectively.

8.5.2 Vision Fusion

As we have two eyes, we should receive two images of an object. Yet we do not see an object double because our brain fuses the two images into one. *Fusion* is the neural process that brings the retinal images in the two eyes to form one single image. It is the fusion process that enables us to have binocular vision, which gives accurate depth perception.

Fusion takes place when the images are from the same object. When the objects are different but their field of view overlaps, suppression, superimposition or binocular (retinal) rivalry may occur to avoid confusion. *Suppression* is the neural process of eliminating one image to prevent confusion. *Superimposition* occurs when one image is presented on top of the other image.

When two very different images are shown to the same retinal regions of the two eyes, our perception settles on one for a few moments, then the other, then the first, and so on, for as long as we care to look. This alternation of perception between the images of the two eyes is referred to as *binocular rivalry* or *retinal rivalry* . This usually occurs when lines with different orientation, size or brightness are presented to the two eyes.

You can experiment the fusion process with the images shown in Figure 8-26 by placing a pen between yourself and the figure. Keep your eyesight on the tip of the pen and move it slowly from the screen towards you. You should see the two images fused together.

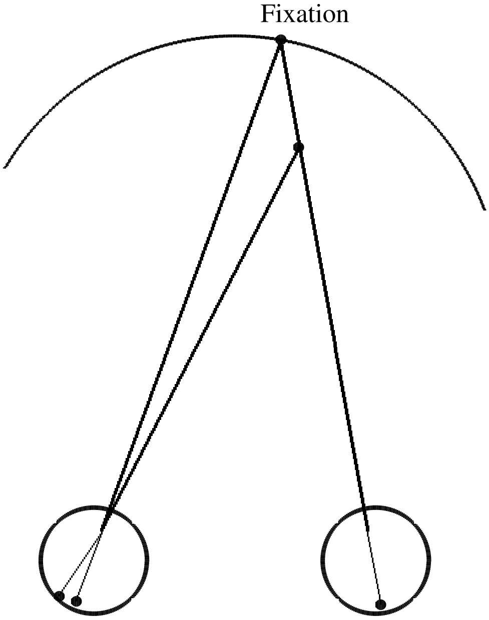

Figure 8-25 Retinal Disparity

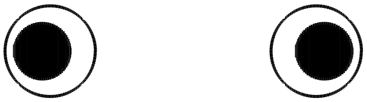

Figure 8-26 Images for Fusion Experiment