Depth Perception

M R Watson and J T Enns, University of British Columbia, Vancouver, BC, Canada

Glossary

Cue The signal or information in a two-dimensional image that can be used to infer or reconstruct the spatial relations in a three-dimensional scene.
Depth Refers to the distance between a viewer and an object as well to the distance between two objects that extend away from a viewer. This is the missing dimension when depicting a three-dimensional scene in a two-dimensional image, whether the image be a picture or the pattern of light projected onto the retina at the back of the eye.
Fovea An area of the retina in the center of each eye in humans (as well as in the eyes of many other sighted animals) containing the most densely packed array of light-sensitive neurons. High-resolution vision is only possible when light from an object falls onto the fovea.
Heuristic A method for solving a problem that begins by first making certain assumptions (e.g., a gradation of shading in an image that runs vertically from light to dark corresponds to a surface with convex curvature, but only on the assumption that light is shining from above the surface and the viewer).
Image A two-dimensional projection of a three-dimensional scene from a specific viewpoint.
Retina(e) The array of light-sensitive neurons at the back of the eye. Light rays passing through the cornea and lens of the eye are focused on the retina and these retinal neurons convert the light into electrochemical signals.
Retinal image The array of light projected onto the retina through the cornea and lens of the eye.

Introduction

Depth perception is the ability of humans and other sighted animals to see objects as having volume (as opposed to seeing flat silhouettes) and to see the relative position of objects in a three-dimensional environment (as opposed to in a two-dimensional picture). This ability is crucial for everyday action. Pushing an elevator button requires information about the button's location relative to our eyes, which acquire the information, and relative to our hand, which must make contact with the button. Picking up a coffee mug requires knowledge of the volume of the handle relative to the cup and to our hand. Thus, depth perception enables interaction with our immediate surroundings. It also grants us information about objects and surfaces that are more distant from us, enabling us to consider future actions. In this article, we review how scientists from several interrelated fields currently understand the problem of depth perception, concentrating on how it is accomplished by humans.

Depth perception is complex and still not completely understood. It begins with the two-dimensional patterns of light that are projected onto the retina, the array of light-sensitive neurons at the back of each eye. When considered in this way – as the problem of recovering the third dimension from a two-dimensional image – depth perception is what scientists refer to as an ill-formed problem. That is, there are infinitely many arrangements of three-dimensional objects that could lead to the two-dimensional array of light that is projected onto the retina. To break this impasse, the visual system uses a wide range of heuristics that are applied to the depth cues that it considers. None of these problem-solving techniques is perfect, but used in combination, they are able to hone in on the likely three-dimensional source of the retinal images.

This article divides depth cues into two broad categories: those cues that are only useful in determining depth for objects relatively near the eyes and those that are reliable over a wide range of distance, both near and far from the eyes. This is an important distinction because humans have evolved a specialized set of mechanisms for near distance vision, capitalizing on the fact that some of the biological apparatus for seeing can alter the nature of the image before it reaches the retina. This includes changing the shape of the lens at the front of the eye to selectively focus the retinal image (accommodation), changing the relative gaze direction of the two eyes to select one of many possible depth planes for sharp focus (vergence), and integrating different images from the two eyes to form a composite binocular image in the brain (stereovision). It is probably not an accident that these structural cues are useful when objects are a short distance from our bodies, enabling us to use these cues when acting rapidly to intercept or avoid objects. Among the cues that are more generally useful, for both near and far vision, we have further distinguished between those that are based on changes in the retinal image that occur over time (depth from motion) and the so-called pictorial or monocular cues (here referred to as static cues to depth).

Specialized Mechanisms for Near Distance

The depth cues that are restricted in their range of distance to within about 3 m (10 ft) from the viewer are known as structural cues, because they involve alterations to the images on the retina by physical changes occurring in a single eye or in the anatomical relations between the two eyes.

Accommodation

The cornea and the lens at the front of each eye are both curved, so as to focus rays of light on the retinal surface at the

back of the eye. In order to form a sharp image at a given distance, the curvature of the eye's lens is adjusted through contraction of cilial muscles that surround the lens, with greater curvature for near objects and reduced curvature for distant ones. This process is called accommodation.

In healthy humans between the ages of 2 months and 45 years, the eye's lens is maximally curved when focusing on objects about 20 cm away (8 in.), and is flattest when focusing on objects at a distance of 3 m or more (10 ft). Neural feedback from the cilial muscles signals the degree of curvature of the lens, which is one indication of the current focal distance of the eye, and hence of the distance of the object the eye is currently focused on, which enables accommodation to be used as a depth cue.

The lenses of newborns are curved to focus on objects about 20 cm (8 in.) away and cannot accommodate for objects at other distances. However, by about 2 months, full accommodation is possible. The material of the lens begins to become more rigid by the age of 20 years and by the time most people enter their mid-40s, the accommodative range of the lens has been sharply reduced. This condition is called presbyopia and is the cause of the widespread use of reading glasses and graduated spectacles in people over the age of 45.

There is some controversy over the utility of this cue in everyday situations because changes in accommodation are rather slow and the neural signal from the cilial muscles is limited in its resolution. A more likely way in which accommodation contributes to the determination of distance is through the amount of blur that is caused in the image when objects are out of focus. Indeed, observers in controlled experiments are able to judge the relative distance of two spots of light in a completely dark room using only one eye, presumably because when one of the spots of light is in focus, the other spot is slightly blurred.

Vergence

The retina contains a small region of tightly packed receptors at the center known as the fovea. In order to see an object at high resolution, the eye needs to be pointed directly at the object such that light reflected from the object falls on the fovea. This is known as fixating the object. Humans typically make three to five fixations every second, moving their eyes to a new locations in order to sample up-to-date, high-quality information about the scene they are viewing. Because the left and right eyes are at different locations in the head, separated by 5–6 cm in an adult, each eye must move differently to fixate the same object.

When the eyes move from fixating a distant object to a closer one, they rotate slightly toward each other (convergence). This increases the angle of convergence between the two eyes, measured as the angle formed at the meeting of two imaginary lines projecting from the pupil of each eye. When the eyes move from fixating a close object to a more distant one, they do the opposite, turning away from each other (divergence). The overall degree of vergence (referring to both convergence and divergence) can be estimated by feedback from the extraocular muscles that control the position of the two eyes, and as such, vergence is a potential cue for determining the absolute distance between the object currently being fixated and the viewer.

The utility of vergence as a general cue to depth is limited, however, by several factors. First, the eyes are already maximally diverged for objects more than 6 m from the viewer, making it potentially useful only as a near distance cue. Second, fairly common eye conditions such as strabismus and amblyopia render vergence ineffective. Third, the signal from the extraocular muscles is quite coarse. Yet, like all the other depth cues, when vergence information is combined with that from other cues, the precision of depth perception improves in a synergistic way. Notably, vergence and accommodation in concert yield more accurate judgments of the distance from the viewer to a single point of light than either cue does alone. Vergence can even be seen working in concert with other depth cues when it is technically unnecessary. For example, research has documented that viewers change their vergence angle when merely viewing pictures. That is, even though all objects in a picture are the same actual distance from the viewer, viewers display greater convergence when looking at objects depicted as closer in the scene than when viewing objects depicted as farther away.

Stereovision

Because our eyes are 5–6 cm apart, the images projected onto each retina are slightly different. This difference in images is a cue to depth called binocular disparity, which enables the experience of depth through the process of stereovision. This process combines corresponding features in each retinal image into a single representation that includes information about distance from the viewer.

We are not normally aware that our eyes contain different images of the same scene, but this can be easily demonstrated. Hold the index finger of each hand in an upright position directly in front of your nose, with one finger about 20 cm away (9 in.) and the other finger about 40 cm away (18 in.). Now focus your eyes on the more distant finger and take turns closing and opening each eye. As you do this, the nearer finger will seem to jump from one side of the farther finger to the other. If you now open both eyes together you should see that there are actually two images of the nearer finger. This is binocular disparity, which enables an accurate perception of depth. The greater the horizontal distance between the corresponding images of the same object in the two eyes (the two images of the closer finger in this demonstration), the greater will be its perceived distance from the object that is currently at the center of the fovea in both eyes (the farther finger).

The positions of an object in the two retinal images are systematically related to the distance of that object from the object that is currently at the center of the two images in each eye. In comparison to the rays of light that project from the fixated object to the center of each retina, light from an object that is closer to the viewer will fall slightly to the right of center in the left eye, and to the left of center in the right eye (this is called crossed disparity). Light from an object that is farther away from the fixated object will do the opposite, falling slightly to the left of center in the left eye, and to the right of center in the right eye (uncrossed disparity). For any object that is fixated, there is an imaginary region of space encircling the viewer at the same distance, called Panum's area. Objects at this distance have no binocular disparity,

meaning that the rays of light projecting from them fall an identical distance from the center of the retina in each eye. As such, these objects also appear to be at the same distance from the viewer as the object currently fixated. Objects outside this region will appear to be nearer or farther, depending on whether they produce crossed disparity (for nearer objects) or uncrossed disparity (for farther objects) in the two eyes. Moreover, the size of the disparity corresponds to an object's relative distance from the fixated object. The process of stereovision, therefore, allows the brain to infer the relative distance of objects on the basis of both the sign (crossed or uncrossed) and the magnitude (size) of the image disparities in the two eyes.

Stereovision can be exploited to create illusions of three-dimensionality, such as seen in Victorian-era stereoscopes, the popular twentieth century Viewmaster series of children's toys, and the glasses worn by audience members at modern three-dimensional films. Though the pictures used in such devices always include depth cues other than binocular disparity, such as occlusion, relative size, and shading (see section on Static Image Cues), it is possible to create a compelling illusion of depth using only changes in disparity, which means that stereovision is a more powerful depth cue that the other structural cues. Bela Julesz invented random dot stereograms at Bell Laboratories in the 1960s to demonstrate this. More recently, the concepts used in making random dot stereograms have been employed to generate the fascinating images known popularly as autostereograms or Magic Eye™ images.

As the name implies, a random dot stereogram appears initially as nothing but a group of dots in a chaotic pattern. However, some of the dots have actually been horizontally displaced relative to one another, such that verging the eyes either in front of or behind the depth of the picture allows an illusion of depth to pop out. When the eyes are focused to the correct distance, each eye's image of the dots is roughly the same, yet some of the corresponding dots in each image are displaced relative to each other. This binocular disparity generates the experience that a subset of the dot pattern has popped into the foreground relative to other regions of the dot pattern that now appear to be in the background.

In addition to demonstrating that stereovision can function independently of other depth cues, random dot stereograms also point to the complexity of the brain's stereovision mechanisms. This is because in order to perceive depth in the pattern of random dots, the brain must somehow know in advance which dots in one retinal image correspond to the same dots in the other retinal image. This is known as the correspondence problem, and like many problems in human vision, it is paradoxically both an ill-formed problem and yet one that the brain seems to solve effortlessly. The fact that it is ill-formed means that in the absence of any information other than that contained in the dot patterns, there are an infinite number of possible ways to align any two retinal images. The fact that the brain solves the problem without effort is interpreted to mean that the brain must be using a priori assumptions about regularities in the environment to solve the problem. A major challenge for vision researchers is to determine what those a priori assumptions are. What is already clear is that the process of stereovision comes to a conclusion more rapidly and more reliably when it is informed by other depth cues, including the monocular cues to depth reviewed later in this entry.

Human infants do not appear to possess functional stereovision at birth, but it develops quite quickly. By the time infants are 6 months of age, most will display stereovision at essentially adult levels. Like the other physiological cues (accommodation and vergence), stereovision is only effectively useful within distances of about 3 m (10 ft) from the viewer. Also, for some of the same reasons mentioned in the discussion of vergence (e.g., conditions of strabismus, amblyopia), between 5 and 10% of the general population does not have usable stereovision because of imbalances in the nature and quality of the information contained in the two eyes.

General Mechanisms for Near and Far Distance

Most of the cues for human depth perception are useful across a wide range of distances. Among these are the cues that derive from motion of objects in the scene and from motion of the viewer, as well as the cues that derive from features in an image seen only by one eye. Here we refer to these single-image cues as static cues, in order to contrast them with the cues available from analyzing motion, but readers should note that they are also called monocular or pictorial cues, since they do not require two eyes and are often used by visual artists.

Depth from Motion

The world is often in motion, giving rise to several rich sources of depth information. For example, when a stationary viewer sees an object in motion, there are systematic changes in what is visible over time. At the leading edge of an object in motion, features of the background that were previously visible will suddenly vanish as the moving object occludes their view (called surface deletion), while at the same time, other features of the background that were previously invisible will suddenly appear (surface accretion). The features of the moving object will remain constantly visible during this time, unless the object is rotating, in which case the systematic deletion and accretion of its features can be used as cues to its three-dimensional shape. Surface deletion and accretion from motion are therefore powerful cues for segregating objects from backgrounds and for determining object volume. These two cues alone are likely the most powerful contributors to the compelling nature of motion pictures, which are not able to benefit from any of the structural cues to depth we have already reviewed, simply because all the information in a motion picture is displayed at the same actual distance from the viewer.

When a viewer moves relative to stationary objects in a scene, and the viewer's eye remains fixated on one object, there are a number of cues to help indicate the relative distance of the various objects in the scene from the viewer. These are collectively referred to as the depth cues from motion parallax. Relative to the object that is at the point of fixation, objects nearer to the observer will move across the retina in the same direction as the observer, while objects further away than the fixated object will move in the opposite direction. Furthermore, the closer an object is to the observer, the faster will its movement be across the retinal image. These changes in the direction and speed of the images of objects across the retina are very effective guides to the relative positions of objects, although less effective for judging the absolute distance of objects.

Motion parallax is not only useful when a viewer is actively and intentionally moving, such as when walking or selectively moving one's head and eyes, but can also be used to accurately judge the relative distance of objects when the viewer is moving passively, as when sitting in a train or in a car watching the world move relative to oneself. However, there are important differences between motion parallax caused by active and passive motion. Most importantly, passive viewers do not generate any muscular feedback from their voluntary actions to refine their passive reception of motion, so the judgments of distance made in such cases are less accurate than when motion is self-generated.

When an object rotates with respect to both its background and a stationary viewer, there is a systematic pattern in the motion of its various parts across the retinal image. These motion patterns contain rich cues for determining the volume of the object. For example, when a car turns toward us, the regions of the retinal image associated with its various parts will move at different speeds, with, for example, the headlight farthest away from us moving at a faster rate than the closer headlight. These differential rates of motion provide the brain with rich information about the relative distance between various object features, thereby allowing the perception of three-dimensional shape to emerge from an analysis of motion patterns alone. This object-relative motion cue was called kinetic depth by Hans Wallach, who first systematically studied the perception of three-dimensional shape that occurs when viewing the shadows of moving objects on a screen, and more recently has been referred to as structure-from-motion by researchers studying the computational complexity of the problem.

When an object moves toward or away from an observer, the size of the region it occupies on the retinal image also increases and decreases proportionately, which is a depth cue from motion known as looming. A particularly important aspect of this depth cue is how symmetrical the changes on each side of the image of the object are. Specifically, a looming image that is expanding symmetrically and rapidly is consistent with an object that will soon hit the viewer on the head. In contrast, a looming image that is expanding asymmetrically is consistent with an object that will move past the viewer without a collision. Human vision is extremely sensitive to looming cues, likely for good evolutionary reasons, and these processes occur largely without awareness.

Research on the development of sensitivity to the entire class of motion cues to depth indicates that these are among the most primary. Newborn infants show sensitivity to the surface accretion and deletion cues of relative motion, to motion parallax, to structure-from-motion, and to looming. Indeed, the newborn startle response to a looming image includes eye blinking, cardiac slowing, and neck and limb stiffening, as though humans are innately programmed to avoid collisions with other objects.

Static Image Cues

Many of the static cues to depth available in static images can be illustrated by considering the beach scene shown in Figure 1.

Researchers have noted that one of the first simplifying assumptions humans make when viewing such a scene is to ignore unlikely possible arrangements of the world. One such unlikely possibility in Figure 1 is that the beach umbrellas are really very different sizes, and have been carefully cut out and positioned into a mosaic pattern, such that they are all at the same distance from the viewer. The visual system ignores this possibility in favor of the more likely interpretation that the umbrellas are all more or less the same size and shape, and that they are positioned at various distances, so that the nearer ones occupy a larger retinal size and occlude our view of parts of the umbrellas that are farther away. This is known as the generic viewpoint assumption, because we assume that the image would not change drastically if our viewpoint on the image were to change slightly. Note, in contrast, that even a small change in viewpoint on this scene would quickly confirm or deny the unlikely mosaic interpretation of the umbrellas. So, in the absence of any evidence to the contrary, humans tend to view scenes, including flat pictures, under the assumption that a small change in viewpoint will not alter the spatial layout of the scene. The generic viewpoint assumption is the basis of a compelling and surprising form of visual art known as anamorphic sidewalk or pavement drawings.

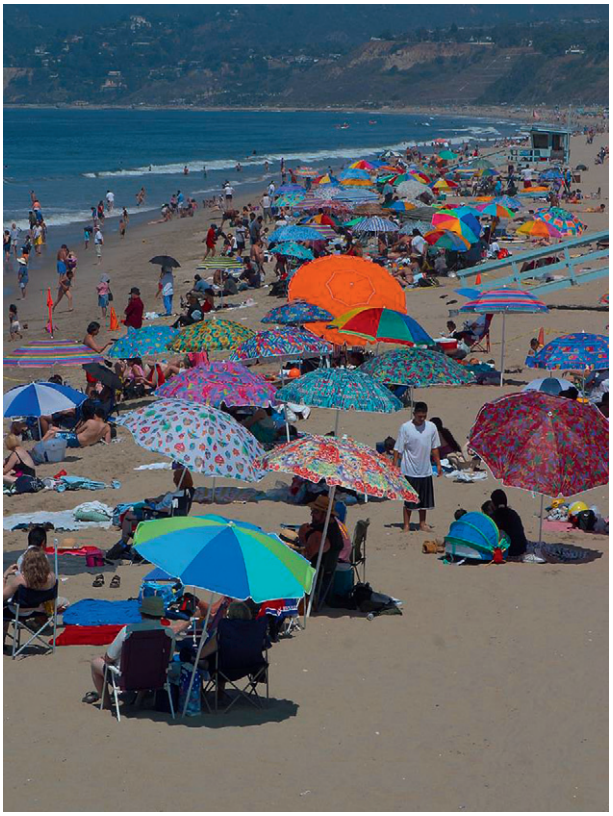

Figure 1 This beach scene from Santa Monica California depicts many of the static cues to depth, including edge intersections, attached and cast shadows, familiar size texture gradients, proximity to the horizon, and aerial perspective. Specific examples of each cue depicted are given in the text.

The static depth cues that must be considered in conjunction with the generic viewpoint assumption will be presented here, for convenience, along a continuum from local to global. By local, we mean that a cue can be considered quite reliably in

isolation from other regions of the image; by global, we mean that a cue is based on information distributed over a large region of the image.

At the extreme local end of this continuum are the junctions that occur when edges intersect with one another in an image. Edge junctions contain some of the most reliable cues to relative depth in an image. For example, in Figure 1, the edges of the large orange umbrella near the center terminate where they meet the edges of the two umbrellas in front of it, forming T-junctions. T-junctions are formed whenever one edge occludes the view of a more distant edge. The edge in front is continuous, whereas the occluded edge terminates at the intersection, making this a reliable local cue to depth. T-junctions are generally considered to be a local cue for what is classically known as the depth cue of occlusion or interposition.

L-junctions are also quite reliable cues to surface relationships in the scene, but must be used in conjunction with a convexity assumption, which is the assumption that, in the absence of any other information, surfaces in the world are more often convex (bulging out) than concave (indented). The convexity assumption applied to L-junctions suggests that the side of the L with the smaller angle belongs to the surface that is nearer to the viewer than the side of the L with the larger angle. L-junctions can be seen in all of the umbrellas in Figure 1, and invariably they occur when an umbrella lies in front of a surface that is further away from the viewer than the umbrella in question.

A second important class of local depth cues concerns the three-edge intersections (i.e., Y- and arrow-junctions) that occur when surfaces are joined together at corners. Corners are extremely useful pieces of information about the three-dimensional shape of objects, and tend to be interpreted in the absence of other information as being convex rather than concave. Note in Figure 1 how the three-dimensional shape of the orange umbrella is revealed by the intersection of its ribs. These edge intersections are ambiguous with regard to convexity–concavity when considered in isolation, but in conjunction with the convexity assumption, they provide rich information about the umbrella shape.

Shadows also convey information about depth. Researchers have found it important to distinguish between attached and cast shadows. In general, the region of a uniformly colored object that is relatively darker than other regions is said to be an attached shadow, so called because this region of the object is being blocked from receiving direct light by other parts of the same object. Attached shadows are particularly useful for determining the three-dimensional shape of objects. For example, in Figure 1, the shadows on the white T-shirt of the young man standing near the front of the picture reveal the shape of his torso. One critical ambiguity that must be resolved in order to use this information to determine the three-dimensional shape of the surfaces is the direction of the light source. All things being equal, the visual system usually assumes that the light in the scene is assumed to be shining from above, which means that if the object has a convex surface, the region of attached shadow lies toward its bottom. However, if the light was actually shining from below, then the same pattern of image shading would indicate a concavity, and the assumption of light shining from above would lead us into error. The interrelatedness of these assumptions in human depth perception can be vigorously exercised by viewing the hollow mask illusion, which occurs when our assumptions of convexity, light generally coming from above, and the familiar shape of faces come into conflict with one another.

Cast shadows help resolve this problem. This can be seen in Figure 1, where the shadows cast by the umbrellas indicate that the sun is almost overhead. Cast shadows provide less information than attached shadows do about the three-dimensional shape of the object casting the shadow, because the shape of the shadow is dependent on both the position of the light source and the shape of the surface they are cast upon. However, they do provide a high-fidelity source of information about the relative positions of the objects in the scene with respect to the light source. The position of the inferred light source can then be used to more accurately judge the shape of objects on the basis of their attached shadows.

Cast shadows are also a rich source of information with regard to relative object position. For example, if a cast shadow borders directly on the object that is casting it, then the casting object is very likely resting on the shadowed surface. On the other hand, if there is a gap between the cast shadow and its casting object, then the casting object must be farther from the surface. The distance between an object and its cast shadow is therefore a cue for their relative positions in space. If this cue is combined with motion, it can be used to create powerful illusions.

Another, somewhat more global, static cue to depth can be derived from comparing the relative retinal size of the images cast by similar objects. In general, it is safe to assume that when two objects of a similar actual size differ in their distance from the viewer, the nearer object will project a larger retinal image. But note that the constraint of objects being of similar actual size rests on its own assumption, namely that the viewer is familiar with the objects and therefore knows that their sizes are similar. When this assumption of familiar size is met, as it is for our perception of the people in Figure 1, then we can reliably see the people of smaller retinal size being farther away in the depicted scene. When some of these interrelated assumptions are violated, as occurs when two objects of similar size are depicted in different apparent locations from the viewer, then even familiar people can appear to vary greatly in their perceived size, as occurs in the Ames Room illusion.

An even more global version of the familiar size cue to depth occurs when linear perspective is used to infer relative distance in a scene. Just as objects of the same size will have smaller retinal sizes as they move further away from the observer, so too will parallel edges in a scene converge toward the horizon as their distance from the viewer is increased. The point at which these lines converge, either within the frame of an image or outside of it, is known as the vanishing point, and corresponds to infinity as far as depth perception is concerned. Note that these lines may be explicit, as occurs when depicting architectural drawings, or they may be implicit, as occurs if we look down a road bordered by trees of a similar height. Here the imaginary lines connecting the tops of the trees converge with the imaginary lines connecting the trunks to the ground at the vanishing point. Linear perspective is the depth cue developed most vigorously in Western art made since the time of the Renaissance until the invention of photography. It is especially effective when used to depict the depth relations among carpentered objects, presumably because these objects tend to have

many parallel edges to calibrate the interpretation of depth. When straight lines cannot be assumed in a carpentered scene, then our tendency to interpret the scene as containing linear perspective can go awry, leading to tantalizing ambiguities and even outright impossibilities, as in the art of M. C. Escher.

When the depth cue of linear perspective is combined with the depth cue of the relative retinal size of familiar objects, it leads to a global depth cue called the texture gradient. A texture refers to any collection of objects in an image, and a texture gradient refers to the pattern of changes in the relative size and spacing of these objects. For example, the beach umbrellas depicted in **Figure 1** form a semiregular pattern, becoming smaller and closer together in the retinal image as they correspond to actual umbrellas that are more distant from the viewer.

Texture gradients inform the viewer about depth through gradual changes in image size and spacing, as seen in the beach umbrellas in **Figure 1**, and the discontinuities in the gradient are themselves informative because they indicate a sudden change in the orientation of a surface relative to the observer. For example, if we are in a room that has been tiled consistently on the floor and wall, we can use discontinuities in the texture gradients to determine changes in surface orientation. These occur where the floor meets the wall and where one wall meets another wall. In the famous study of human infant's sensitivity to a visual cliff by Eleanor Gibson and Richard Walk, only a texture gradient was used to signal to the infants that the surface beneath the glass floor they were crawling on consisted of surfaces at two different distances.

At the most global end of the continuum are several static depth cues that require a comparative analysis of almost the entire visual field. Proximity to the horizon (also known as relative height or height in the plane) refers to the observation that objects that are nearer to the horizon are generally seen as being farther away from the viewer than are objects that are farther away from the horizon. This means that below the skyline, more distant objects tend to be higher in the retinal image. Thus, in **Figure 1**, the people in the water are seen as more distant than the man in the white T-shirt. Above the horizon, these relationships are reversed, with nearer objects in the scene being higher and more distant objects being closer.

Finally, aerial perspective is a global depth cue that is based on relative differences in contrast and color rather than on size or spatial position. When light travels through the air, it is either scattered or absorbed by various particles. This means that the light from distant objects travels through more air before reaching our eyes than the light reflected from nearby objects, with the consequence that less of a distant object's reflected light reaches the eye. As such, distant objects will tend to be more blurred than nearby objects (relative contrast), and they will tend to be darker (relative brightness). They will also have a bluish tinge (relative hue), because air molecules scatter shortwave blue light to a greater extent than light of longer wavelengths, and some of the blue light from the sun is scattered toward the viewer. This can be seen in the hills at the top of **Figure 1**.

Not surprisingly, given the complexity and interrelatedness of the static cues for depth, these cues tend to appear more slowly in human development than the structural and motion-based cues. Sensitivity to edge intersections develops relatively early in life, with research suggesting that by 3–4 months infants are responding to T-junctions in an adult-like manner. However, it is not until 6–7 months that most infants can respond reliably to linear perspective, texture gradients, familiar size, or shading. The reliable use of cast shadows may not be fully developed until 3 or more years of age.

Depth Perception in the Brain

There is currently much scientific interest in understanding how the various cues for depth are processed and combined by the neurons of the brain. The cue that has been studied most thoroughly in this regard is binocular disparity, which, as discussed in an earlier section, is critical for the computation of stereovision. Neurons that are tuned to specific binocular disparities have been found in numerous regions of the visually sensitive cortex of primates (including humans). Roughly speaking, the further removed these neurons are from the eye (by virtue of the number of synapses that must be crossed for information to reach them), the more accurate and refined is the stereovision they exhibit. Neurons in area V1, the first stage of cortical visual processing, are sensitive to stimuli in the two eyes that satisfy very loose criteria of similar disparity. Neurons in the next cortical region, area V2, function under stricter criteria, implying that more of the correspondence problem has been solved at this point. In keeping with this trend of increasing sophistication of processing as we move away from the eyes, neurons in higher cortical regions are also specialized to different *kinds* of disparity. For example, one class of neurons may respond only to abrupt changes in depth, whereas other neurons may register more continuously graded changes.

Conclusion

This review of depth perception in humans has emphasized what is currently known about the way various physical factors – both in the structures of the eye(s) and in features of the image – are used by the brain to achieve the experience of depth perception. The review organized depth cues into two broad categories: those specialized for depth at near distances, where direct bodily interaction with other objects is most likely, and those cues that work at any distance from the eyes. The near cues included changing the shape of the eye's lens (accommodation), changing the relative gaze direction of the two eyes (vergence), and integrating disparate images in the two eyes (stereovision). Among the cues more generally useful for both near and far vision, the review further distinguished between those that involve changes in the retinal image over time (depth from motion) and those that are effective when they appear in a picture (static image cues).

See also: Visual Motion Perception; Visual Perception.

Further Reading

Braunstein ML (1976) *Depth Perception Though Motion*. New York: Academic Press.
Coren S, Ward LM, and Enns JT (2004) *Sensation and Perception*, 6th edn. New York: Wiley.
Enns JT (2004) *The Thinking Eye, the Seeing Brain*. New York: WW Norton.

Howard IP and Rogers B (2002) *Seeing in Depth*, vols. 1 & 2. New York: Oxford University Press.
Julesz B (1971) *Foundations of Cyclopean Perception*. Oxford, England: University of Chicago Press.
Parker AJ (2007) Binocular depth perception and the cerebral cortex. *Nature Reviews Neuroscience* 8: 379–391.
Wallach H and O'Connell DN (1953) The kinetic depth effect. *Journal of Experimental Psychology* 45: 205–217.
Wolfe JM (ed.) (2007) *Sensation and Perception*. Sunderland: Sinauer.

Relevant Websites

http://hyperphysics.phy-astr.gsu.edu/HBASE/vision/accom.html – Accommodation.
http://psych.hanover.edu/krantz/art/aerial.html – Aerial Perspective.
http://psych.hanover.edu/krantz/MotionParallax/MotionParallax.html – Depth from Motion.
http://www.michaelbach.de/ot/mot_ske/index.html – Depth from Motion.
http://www.richardgregory.org/experiments/index.htm – Depth from Shading.
http://vision.psych.umn.edu/users/kersten/kersten-lab/demos/BallInaBox.mov – Depth from Shading.
http://en.wikipedia.org/wiki/Depth_perception – Depth Perception in General.
http://psych.hanover.edu/Krantz/art/cues.html – Depth Perception in General.
http://www.sinauer.com/wolfe/chap6/startF.htm – Depth Perception in General.
http://www.european-street-painting.com – Generic Viewpoint Assumption.
http://www.ski.org/CWTyler_lab/CWTyler/ArtInvestigations/PerspectiveHistory/Perspective.BriefHistory.html – Linear Perspective.
http://psych.hanover.edu/KRANTZ/art/rel_hgt.html – Proximity to the Horizon.
http://www.psychologie.tu-dresden.de/i1/kaw/diversesMaterial/www.illusionworks.com/html/ames_room.html – Retinal and Familiar Size.
www.mcescher.com – Static Image Cues.
http://www.yorku.ca/eye/disparit.htm – Stereovision.
http://cprr.org/Museum/Ephemera/Stereo-Viewers.html – Stereovision.
http://www.vmresource.com/ – Stereovision.
http://www.magiceye.com – Stereovision.
http://psych.hanover.edu/krantz/art/texture.html – Texture Gradients.
http://library.thinkquest.org/27066/depth/nlotherdepth.html – Vergence.