INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume3, Issue1)
Available online at: www.ijariit.com

Human Depth Perception

Ajit Kumar Sharma
Department of BCA,
R.N.College, Hajipur (Vaishali),Bihar
ajit_rnc@yahoo.com

Kiran Kumari
Department of Physics
R.N.College, Hajipur (Vaishali),Bihar
drkirankumari97@gmail.com

***Abstract:** we introduce the perceptual issues relevant to seeing three dimensions in digital imagery. Technological constraints like limited field-of-view and spatial resolution prevent the display of images that match the real world in all respects. Therefore, only some elements of real world depth perception are utilized when viewing 3D CGI. Depth Cue Theory is the main theory of depth perception. It states that different sources of information, or depth cues, combine to give a viewer the 3D layout of a scene. Alternatively, the Ecological Theory takes a generalized approach to depth perception. It states that the HVS relies on more than the image on the retina; it requires an examination of the entire state of the viewer and their surroundings (i.e., the context of viewing). In this paper, we rely on Depth Cue Theory, although we acknowledge the importance of visual context where appropriate. As seen later, the type of visual environment and the viewer's task play a significant part in the effectiveness of a 3D VDS. Both theories assert that there are some basic sources of information about 3D layout. These are generally divided into three types: pictorial, coulometer and stereo depth cues. The perceptual process by which these cues combine to form a sense of depth is a complicated and outdebated issue. Different approaches to measuring the ability to perceive depth have also been posited. We discuss these issues with respect to CGI.*

***Keywords:** Simulators, CGI, Stereo Viewing.*

I. INTRODUCTION

Over the past forty years, advances in display and computing technology have revolutionized the interface between man and machine. Now that people can interact with rich, realistic, 3D graphics with relatively low cost equipment, the time has come to focus on designing our systems so that we maximize their capabilities in the ways most effective for the user.

Man-machine interfaces found in simulator and Teleoperation systems have laid the groundwork for completely computer-generated or virtual environments. *Simulators* are training systems that display computer-generated scenes based on real-world situations. *Teleoperation systems* extend a person's ability to sense and manipulate the world to a remote location. The control and display devices in these systems are often computer-controlled. *Virtual environments (VEs)* are computer generated experiences that may seem real but are not required to match any of the rules of the real world. Virtual environments are synthetic sensory experiences that communicate physical and abstract components to a human operator or participant. The synthetic sensory experience is generated by a computer system that one day may present an interface to the human sensory systems that is indistinguishable from the real physical world.

VEs, like many simulators and teleoperation systems, rely on the visual display systems (VDSs) ability to match the users sensory channels. In VEs, the visual channel is often the most prominent. Therefore, improving the quality and capabilities of the VDSs used in VEs is vital. Unfortunately, while advances in processing power have occurred at an exponential rate in recent years, advances in visual display technology have not. This has led to a variety of problems in how to effectively and efficiently present realistic 3D, computer-generated imagery (CGI).

This dissertation studies how visual display systems can better meet the requirements of the human visual system (HVS). Specifically, we argue that:

1) HVS requirements are a function of the type of task performed in a VE.
2) There is a relationship between task performance and VDS characteristics.
3) Task-centered analysis can lead to new, more efficient techniques for improving the design and display of 3D imagery.

Pictorial or *monocular depth cues* are 2D sources of information that the visual system interprets as three-dimensional. Because pictorial cues are 2D, the depth information they present may be ambiguous. Many common optical illusions are based on these ambiguities [Gillam 1980][1]. Despite the potential for ambiguity, combining many pictorial depth cues produces a powerful sense of three dimensionality.

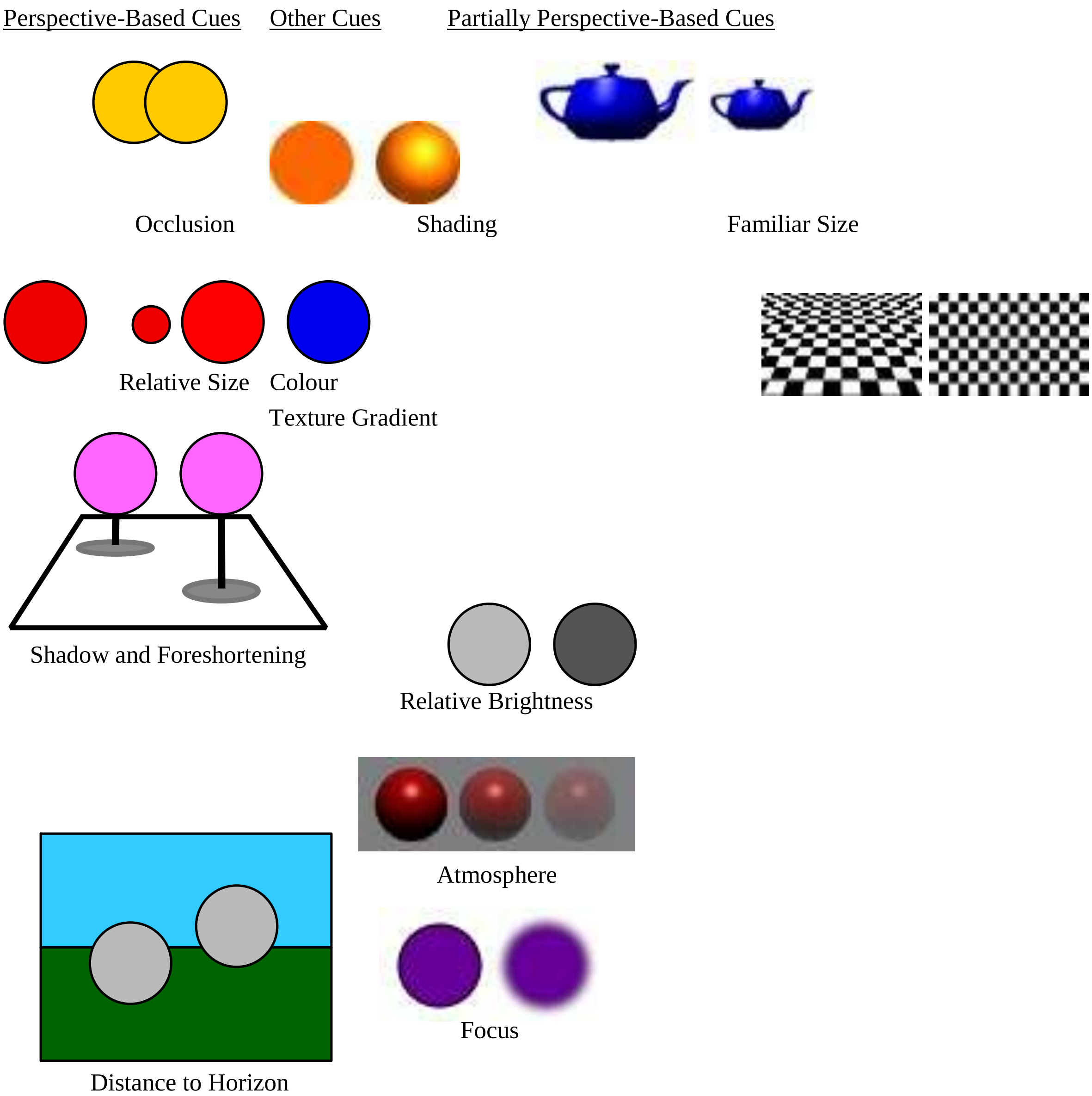

Figure 2.1: Pictorial depth cues in static images classified by their reliance on perspective geometry.

Figure 2.1 shows one classification of static pictorial depth cues, although different taxonomies are often used [Cutting & Vishton 1995; Goldstein 1989][2]; Sedgwick 1980][3]. In this thesis, we consider all cues whose magnitude is governed by the geometry of perspective projection to be *perspective-based cues*. For example, the amount of one object occludes another is determined by the location of the viewer relative to the objects, and thus the perspective geometry of the scene. Pictorial cues in moving images.

Art History describes the use and development of many of these cues. Since linear perspective was rediscovered in the Renaissance by Brunelleschi, Drer and Alberti, these cues have been used extensively by artists [Gombrich 1969; Pizlo & Scheessele 1998][4].

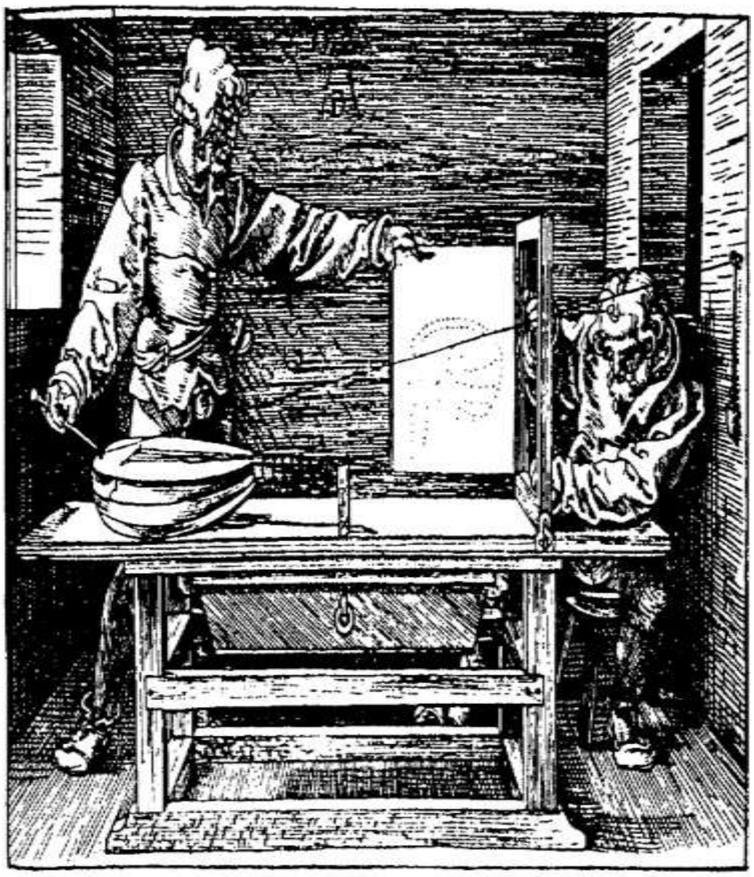

Figure 2.2: Albrecht Drers 1525 woodcut Drawing a Lute showing the construction of a perspective projection. Copyright New York City Public Library.

The construction of linear perspective drawings is a well-documented technique taught to architects and artists [Bartschi 1981][5]. The ambiguity of perspective can be seen in M.C. Eschers renditions of impossible scenes or Ames laboratory trompe lil [Gombrich 1969; Ittleson 1952][6].

Figure 2.3: M.C. Eschers 1938 woodcut Waterfall demonstrating ambiguity in pictorial depth cues(7). [Copyright 1988 M.C. Escher heirs/Cordon Art, Baarn, The Netherlands].

This ambiguity leads to errors in the judgment of size and distance within a scene (9) [Baird 1970]. Similar problems are found in CGI, but are often attributed to restricted viewing angles (10) [Kline & Witmer 1996].

Pictorial cues are often used to convey depth in CGI because most commonly available VDSs are only capable of presenting 2D images. However, even a simple 2D VDS is capable of presenting a compelling 3D image by using many redundant pictorial cues and combining them with depth information from object or viewer motion.
Producing a variety of detailed pictorial depth cues is often computationally expensive. To correctly compute the shading, colour and lighting for a complex scene and thus accurately present depth cues derived from these features is difficult to do in real time. Even with specialized hardware systems, rendering a large number of polygons, using only perspective depth information can be computationally intractable.

As a result, real-time applications often forego the level of realism attainable with algorithms that are more complex. In some cases, this means that depth cues are presented less accurately. For example, wireframe models may be substituted for shaded models to improve performance, but doing so removes occlusion depth cues. Alternatively, texture maps may be reduced in size (and thus resolution), which results in degraded texture gradient depth cues. To help designers make these kinds of choices, the relative effectiveness of depth cues in CGI has been investigated (11) [Surdick et al. 1994; Wanger, Ferwerda & Greenberg 1992]. Among pictorial depth cues, linear perspective is widely regarded as one of the most effective sources of depth information in 3D CGI (12) [Hone & Davies 1995].

Oculomotor depth cues include *convergence* and *accommodation*. Convergence is the rotation of the eyes towards a single location in space. Accommodation is the focusing of the eyes at a particular distance. Because these cues are dependent on each other and on binocular depth cues, their effect on depth perception is difficult to measure [Gillam 1995]. Although including oculomotor cues is considered important for immersive viewing, compelling scenes can be constructed without these depth cues, at the cost of producing visual after-effects(12) [Rushton & Wann 1993]. VDSs using stereo imagery also have to account for problems related to oculomotor cues.

Stereopsis, or the use of the *binocular disparity depth cue*, is the process by which the angular disparity between the images in the left and right eye is used to compute the depth of points within an image.

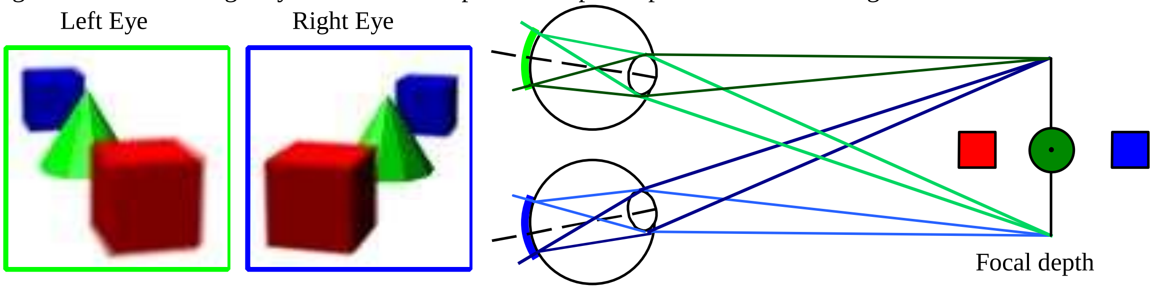

Figure 2.4: Binocular and Oculomotor depth cues. The images on the left show the left and

Right eye views resulting from a binocular view of the scene shown in plan view on the right. Oculomotor information results in the depth of focus shown in the images, where the green cone is in focus and the red and blue cubes are not.

Sir David Brewster has been credited with encouraging popular interest in stereo depth cues with the development of the stereoscope (13)[1856]. In modern day immersive systems, stereo display is believed to contribute to a sense of presence. Despite the continuing popularity of stereo presentation, its use in 3D CGI is often questioned (14) [Hsu et al. 1994]. As a result, binocular disparity has been studied more than any other depth cue with respect to CGI.

II. DEPTH FROM MOTION

Motion cues to depth provide information about the location, velocity, acceleration and direction of movement of either the viewer or an object. The Ecological Theory of vision argues that because a human viewer is always experiencing some kind of motion, perception is best studied in terms of the changing information in the optic array (i.e., texture gradients and flow patterns)(15) [Gibson 1986]. Since most CGI is defined geometrically, describing non-geometrical depth information can be awkward. In most situations in this dissertation, we simplify by assuming a static viewpoint.

Motion of stereo information is especially important since the visual system is more sensitive to binocular disparity when it is changing ((16) [Yeh 1993]. The literature suggests some general characteristics of the depth perception derived from motion cues in experimental settings. In some cases, this research can be extended to situations that are more practical. To better understand the temporal aspects of depth perception and sampling in 3D CGI, we examine perspective and general, the more cues presented, the better the sense of depth (Figure 2.5). Geometric enhancements can reduce the ambiguity of pictorial depth cues [Ellis 1993]. However, the best way to disambiguate pictorial depth cues is to present stereo depth information. stereo cues in tasks requiring accurate perception of motion.

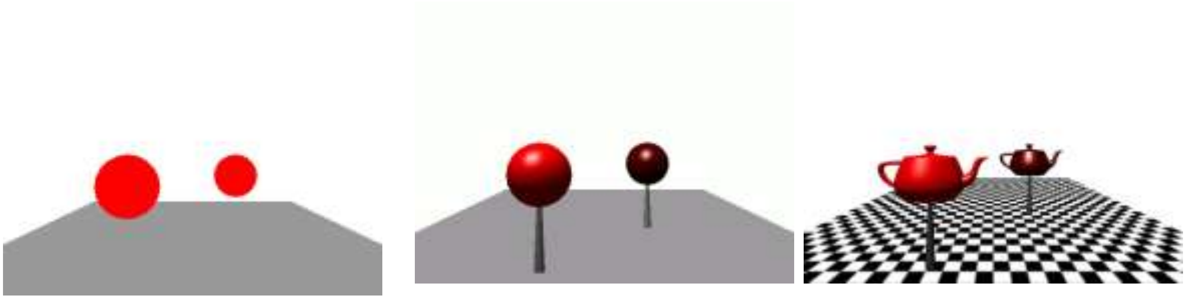

Figure 2.5: Adding depth cues improves the sense of depth in a pictorial image, as shown in figure

III. COMBINATION AND APPLICATION OF DEPTH CUES

All the depth cues we are discussed combined by the HVS to give a sense of 3D layout. In the increasing sense of three-dimensionality from the leftmost image to the rightmost image. Some cues dominate others in certain situations. For example, a person threading a needle primarily uses stereo cues to determine the location of the end of the thread and the eye of the needle, and usually brings the objects close to the eyes to increase the accuracy of stereo and oculomotor cues. However, a submarine pilot is unlikely to use stereo or oculomotor cues to determine the distance to a far-off buoy, instead relying on multiple pictorial depth cues(13) [Pfautz 1996]. An important criterion for the dominance of one cue over another is the distance from the viewer to the objects of interest

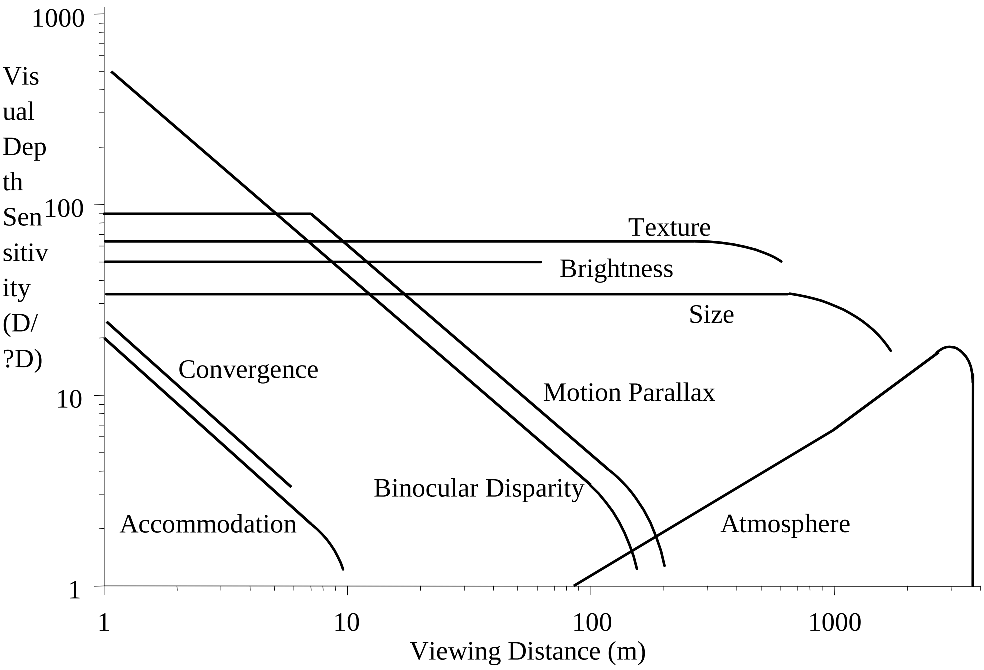

Figure 2.6: The effectiveness of depth cues as a function of distance.

The rank linear perspective as the most effective across all viewing zones. They also note that binocular and oculomotor cues decrease in value with increased viewing distance.

In 3D CGI, increasing the displayed depth (i.e., the depth of the object according to the various depth cues shown in the scene) decreases the effectiveness of some depth cues. Linear perspective and stereo cues are among the most effective over a range of displayed depths. Other cues, like luminance or contrast, have comparatively little effect. Although other depth cues are important in some visual contexts, this dissertation only considers linear perspective and stereo depth cues in 3D CGI.

We will only treat the relationship between stereo and perspective depth cues in this paper. Estimating time-to-collision can be performed more effectively as more depth cues are provided, regardless of potential conflicts. Dynamic stereo information has been shown to improve the ability to accurately track objects in 3D CGI when combined with potentially conflicting perspective cues. As with static depth cues, the accuracy of judgments about an objects motion will increase with the number of different depth cues describing that motion.

Applying and combining depth cues to create a 3D scene is a complex process. A designer has to consider all of the following:

- Increasing the number of pictorial cues decreases a scenes ambiguity.
- Binocular disparity cues may disambiguate pictorial cues, but are difficult to present without conflicts with accommodation and convergence.
- Different cues are more effective at different distances.
- And, most importantly, the value of a depth cue varies with the type of task.

To evaluate depth cues effectiveness, we need a measure of accuracy of depth perception. In 2D, *spatial acuity* refers to the HVS ability to discriminate points and objects. Spatial acuity varies with the type of test, viewing conditions and a number of other factors . The most common spatial acuity test is the Snellen test, where different sizes of letters must be identified and standard acuity (for a 20 year old at 20 feet) is expressed as 20/20 (equivalent to 1' of visual angle in other tests). A viewer with 20/300 vision would be able to see details at 20 feet that a normal 20/20 viewer could see at 300 feet. In this dissertation, we will use degrees of visual angle rather than Snellen acuity, as it is more common in Psychophysics and Human Factors literature.

Depth acuity is the ability to discriminate points in depth. Depth acuity is tested in a variety of ways . *Stereo acuity* is the ability to differentiate two points in depth using only binocular disparity. Stereo acuity ranges as a function of viewing distance, from 0.05mm at 0.25m to 550mm at 25m. In terms of visual angle, standard stereo acuity means a viewer can discriminate disparity differences of around 2' of arc.

The Howard-Doleman apparatus is the canonical method for measuring both stereo and depth acuity. A subject views two cylindrical posts through a rectangular aperture and adjusts distance of one to match the other . With this method, depth cues can be added or removed to measure their individual effects on acuity.

Variations on the Howard-Doleman apparatus have also been used to evaluate the accuracy of depth judgements. A peg-in-hole manipulation task is a common variation [Matsunaga, Shidoji & Matsubara 1999], as are 3D tracking tasks [Zhai, Milgram & Rastogi 1997]. Like spatial acuity, the thresholds found for depth acuity are often a function of the type of experimental task [Surdick et al. 1994; Wanger, Ferwerda & Greenberg 1992].

Perhaps the greatest difficulty with measuring depth acuity is effectively dealing with multiple depth cues. The inherent ambiguity of many pictorial cues means experiment design can be frustrating. Testing cues individually may reveal biases not seen in the presence of other cues. As a result, experiments on artificial scenes with overly sparse depth information may not generalise well.

CONCLUSIONS

We discussed the different depth cues and their application in CGI. In part because of their ubiquity in 3D CGI and in part because previous work has emphasized their importance, we will consider the effects of sampling on only linear perspective and stereo depth cues in this paper

REFRENCES

1. The ability to fuse two images into a single image is described in terms of the *horopter*, a circle in space defined by points that fall onto corresponding points on the two retinae. Points that lie on the horopter will be fused into a single image [Graham 1951].
2. *Panum's area* is the range in front and behind the horopter where single images can be seen [Buser & Imbert 1992]. This range is mainly a function of the viewing distance and is important in the design of stereoscopic display systems (Chapter 3).
3. Binocular vision can only occur where the field-of-views of the two eyes overlap. The horizontal binocular visual field is about 120° out of a possible 200° (Figure 3.1).
4. Stereopsis plays an important role in fine discrimination of objects in the near- and mid-fields but has a diminished role for objects more than ten metres from the viewpoint [Nagata 1993].
5. Accommodation results in some convergence [Gillam 1995].
6. Some stereo VDSs present binocular information in conflict with accommodation and convergence information [Wann, Rushton & Mon-Williams 1995].
7. Relative size cues may interfere with motion perspective cues [Delucia 1995].
8. Stereo viewing of pictorial depth cues results in conflicting depth information, thus decreasing the sense of depth [Gombrich 1969].
9. *Motion parallax*: objects moving parallel to the viewer move faster in the near visual field and slower in the far field.
10. *Deletion/accretion* or *kinetic occlusion*: change in the amount one object obstructs the view of another [Goldstein 1989]. • *Motion perspective*: movement of points in space according to the laws of linear perspective.
11. *Familiar speed*: perception of layout given a velocity that is familiar to the viewer (e.g., the second hand on a watch or a person walking) [Hochberg 1986].

