Forthcoming in *ACM Transactions on Computer-Human Interaction*.

The Relative Contributions of Stereo, Lighting and Background Scenes in Promoting 3D Depth Visualization[1]

GEOFFREY S. HUBONA[2]
Virginia Commonwealth University, School of Business, Department of Information Systems, 1015 Floyd Avenue, P.O. Box 844000, Richmond, VA 23284-4000
PHILIP N. WHEELER
AppNet Inc., 1100 West St., Laurel, MD 20707
GREGORY W. SHIRAH and MATTHEW BRANDT
NASA Goddard Space Flight Center, Codes 935 and 588, Greenbelt, MD 20771

More powerful contemporary computer hardware has enabled the development and exploration of a wide variety of techniques to depict spatial characteristics of computer-generated objects in three dimensional (3D) space. Particularly, the role of stereoscopic viewing and the use of object motion to reflect the position and size of objects in 3D space have been extensively studied. However, the effective use of computer-rendered object shadows to provide spatial information about the relative position and size of objects in virtual space has not. Subjects perform two tasks with 3D geometric patterns of objects presented on a computer screen: (1) positioning the object to complete a symmetrical geometric figure; and (2) resizing the object to match the size of other objects. Performance accuracy and speed are recorded under the following conditions: (1) objects casting shadows on and off; (2) shadows from one or two light sources (nested within the shadows on condition); (3) stereoscopic and monoscopic viewing; and (4) different scene backgrounds: flat plane (i.e. floor); 'stair-step' floor with no walls; and floor with walls (i.e. room). The use of object shadows as depth cues enhances the accuracy (but not the speed) of object positioning, but does not enhance either the accuracy or the speed of object resizing. Moreover, the object shadows are not as effective as stereoscopic viewing in facilitating both positioning and resizing task performances. Furthermore, task performances degrade with the stair-step scene background, and when the number of shadowing light sources increases from one to two.

Categories and Subject Descriptors: H.1 [**Models and Principles**]: User/Machine Systems; H.5 [**Information Interfaces and Presentation**]: User Interfaces; I.3 [**Computer Graphics**]: Three-Dimensional Graphics and Realism

General Terms: Experimentation, Human Factors

Additional Key Words and Phrases: 3D user interfaces, shadows, stereoscopic viewing, depth perception, cue theory

1. INTRODUCTION

There has been a dramatic increase in the performance capabilities of personal computers and modern workstations in recent years. The increasing power of computing machinery has enabled the proliferation of three-dimensional (3D) user interface techniques for a variety of business, entertainment, and scientific applications. For example, a number of 3D visualization prototypes are under development in the Advanced Visual Tools and Architectures[3] (AVATAR) and in the

---

[1] This research was sponsored by the Advanced Visual Tools and Architectures in Code 588 and by the Scientific Visualization Studio in Code 935 of the Goddard Space Flight Center (GSFC) of the National Aeronautics and Space Administration (NASA) in Greenbelt, Maryland, USA.

[2] Authors' email addresses: gshubona@atlas.vcu.edu, pwheeler@appnet.com, greg.shirah@gsfc.nasa.gov, and matt.brandt@gsfc.nasa.gov.

[3] See http://nipper.gsfc.nasa.gov/vetdocs/VETHomePage.html.

Scientific Visualization Studio[4] (SVS) at the National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC).

However, a key challenge facing 3D interface designers is to develop effective techniques to depict objects in 3D space on a physical medium that is inherently two-dimensional (2D): a flat computer screen. It is especially important to depict spatial relationships among objects in 3D space, particularly with respect to depth (e.g. the *z* dimension) so that users can locate and manipulate these objects. To this end, interface designers use various depth cues to make objects 'appear' 3D on a 2D screen. These techniques are based on psychological research about human perception and the cognitive processing of visual information.

The visual cues that enable humans to perceive depth have been extensively documented. However, the use of these cues, and particularly, the combinations of these cues, to effectively convey depth information in computer-generated scenes remains an ongoing topic of research. Often depth cues are placed in two categories: (1) primary cues including binocular disparity, convergence and accommodation; and (2) secondary cues that include perspective and elevation, size, texture, shading and shadow, motion, reference frames, and others (see Kelsey [1993] for a complete discussion of these primary and secondary depth cues).

Binocular disparity is a primary, physiological characteristic that enables the stereoscopic viewing of objects within a limited distance. Stereo viewing facilitates spatial tasks by providing powerful depth information about the relative sizes and positions of objects in space. However, there are practical limitations with respect to the use of stereo vision-enabling hardware and software in computer applications. It is relatively expensive, requires the concomitant use of often cumbersome head-mounted apparatuses, and it is not supported by most existing software. For these reasons, it is useful to research visual techniques that may serve as practical alternatives to stereo viewing in conveying depth information.

The use of real time, moving object shadows to convey depth information about the corresponding objects in 3D space is a technique that merits additional study. In particular, the relative contributions of moving, real-time object shadows in facilitating performance with interactive, spatial tasks has not been fully examined, and deserves additional study. In part, this is because contemporary hardware has just recently become powerful enough to support practical applications involving the computation and rendering of real time shadows for user-controlled, moving objects.

Moreover, the influence of the *configuration* of the scene background has not been widely examined and merits additional study. To this end, we utilize three background configuration metaphors that exist in the 'real world': a flat plane, a room, and a stair-step. The room and stair-step backgrounds are more complex configurations than the flat plane, but one might argue that *more* depth information can be gleaned by watching the shadows 'cascade' along the intersections of the walls or down the steps. We intuitively feel that this aspect is of interest to researchers in this area.

The purpose of this study is to investigate how the presence (or absence) of shadows, created by one or two light sources on objects viewed in scenes with different backgrounds, provides depth information about the relative placement and size of objects in 3D space. It is also of interest to compare the relative contributions of object shadows with stereoscopic viewing in benefiting interactive task performances involving the positioning and resizing of objects in 3D space. This article reviews the more powerful 3D depth cue display techniques and then presents a formal experimental study that examines the power of cast shadows in providing depth

---

[4] See http://svs.gsfc.nasa.gov.

information so as to facilitate 3D object positioning and resizing tasks. The findings are discussed with respect to theory and future research.

2. THEORY AND BACKGROUND

2.1 Cue Theory

Theoretical studies on the mental representations of visual graphic images have examined the perception of multidimensional stimuli. A *multidimensional* computer-generated image is one that combines several realism (or depth) cues (Lockhead [1972] called them ‘stimulus dimensions’), such as the number of light sources, shadows, surface shading, texture, color, and so forth. An important question is whether these stimulus dimensions are perceived as separate elements, or whether they are perceived as belonging to an integrated, unitary whole object [Garner and Felfoldy 1970; Foard and Nelson 1984].

Cue theory suggests that the visual system implicitly computes the distances of environmental objects on the basis of information about the posture of the eyes and about patterns of light projected onto the retinas. While it is important to understand how individual depth cues contribute to the holistic perception and interpretation of depth information in computer-rendered scenes, it is equally important to understand how the visual system integrates this information to foster a singular, stable perception of three-dimensionality. It is also relevant to consider whether the depth information provided by the different cues is complementary or conflicting in nature. Several models have been proposed, including the *additive* and *multiplicative* models; the *vetoing hypothesis*; and *weak fusion* and *strong fusion*.

Bruno and Cutting [1988] conducted experiments in which depth information was provided by several non-conflicting pictorial cues, including relative familiar size, height, occlusion and motion parallax. Although these cues provided redundant depth information, subjects always perceived more depth in displays characterized by the presence of more cues. They concluded that depth information from various cues is combined in an *weighted additive* fashion. That is, the combined effects of multiple depth cues can be determined as the weighted linear summation of the information provided by each individual cue, with individual numerical weightings, or strengths, ascribed to each separate cue. In contrast are *multiplicative* models which suggest that the synergistic informational effects of multiple depth cues are combined in a more complex fashion. Multiplicative models posit that depth cues interact and can foster either a ‘greater-than’ or ‘less-than’ additive effect. Sollenberger [1993] reported evidence for the multiplicative model in experiments combining motion parallax and binocular disparity depth cues.[5]

Johnston et al. [1993] suggested a *vetoing* mechanism in situations where strongly conflicting information is provided by different cues. They suggested that in these situations, the stronger, or more dominant depth cue, simply overrides the effect of the weaker cue such that the combined perception of depth is equivalent to that perception of depth resulting from the stronger cue alone. Johnston et al. [1993] cited experiments by Bülthoff and Mallot [1988] in which the depth information provided by the zero retinal disparity cue vetoed depth information provided by shading, resulting in the perception of a flat surface.

Johnston et al. [1993] attributed the *weak fusion* and *strong fusion* mechanisms for cue combination to Clark and Yuille [1990]. The *weak fusion* model, similar to the weighted additive

---

[5] For a complete discussion of additive and multiplicative approaches to depth cue information integration, please see Cutting et al. [1992].

model, suggests that depth information is processed separately by each cue, and then combined in a weighted linear fashion to compute the overall effect. Young et al. [1993] proposed a *modified weak fusion* model in which independent depth estimates formed on the basis of separate depth cues are weighted, and then additively combined by a factor reflecting their "apparent reliability" in that particular visual context. In contrast to *weak fusion* models, the *strong fusion* model suggests that depth cues interact non-linearly. For example, one cue may work to disambiguate information, enabling depth information to be derived from another cue.

2.2 Dominant Depth Cues

Stereopsis is a powerful depth cue, particularly for objects that are relatively close to the viewer [Yeh 1993]. However, we perceive depth, and the relative positions of objects in space, for even the most distant objects. Thus, stereopsis is not the only mechanism for perceiving depth. Numerous studies have investigated human performance capabilities with stereoscopic user interfaces in various task domains, including: cockpit situational awareness [Bridges and Reising 1987; Andre et al. 1990; Reising and Mazur 1990; Yeh and Silverstein 1990]; the viewing, manipulation, grasping and/or recognition of object images [Gallimore and Brown 1993; Hubona et al. 1997; Wickens et al. 1994; Brown and Gallimore 1995; Ware and Franck 1996; Zhai et al. 1996]; relative depth perception [Reinhart 1990]; and medicine [Sollenberger and Milgram 1993]. Studies have shown stereo viewing to be a powerful technique for displaying depth information [McAllister 1993; Wickens et al. 1989] and stereo viewing has been used as a control condition, or baseline, for studying other 3D display techniques [Arthur et al. 1993]. Many studies do indicate the benefits of stereoscopic viewing in perceiving, recognizing, grasping and/or understanding object shapes [Hubona et al. 1997; McWhorter et al. 1991; Sollenberger and Milgram 1993; Brown and Gallimore 1995; Ware and Franck 1996; Zhai et al. 1996], although some of the studies do not support the superiority of stereopsis [Gallimore and Brown 1993] as a depth cue and others suggest that these benefits are task specific [Liu and Wickens 1992b; Wickens et al. 1994]. However, stereopsis does provide depth cues about object shape that are absent without stereopsis. The general consensus is that stereopsis is a powerful (perhaps dominant) cue for providing the viewer information about the relative location, size, shape, and orientation of objects in 3D space.

Another powerful depth cue is *motion*. Everyday perceptual experiences occur within a context of nested motions. Moving eyes, and moving objects, provide powerful perceptual cues about the environmental and spatial properties of perceived objects. Sollenberger and Milgram [1993] demonstrated the utility of *motion parallax* in graphically visualizing complex, simulated blood vessel structures in the brain. Overbeeke and Stratmann [1988] and Smets [1992] demonstrated that the motion cue can be introduced through the observer's own head movement. Arthur et al. [1993] and Ware and Arthur [1993] found that the use of head-tracking displays (to introduce the motion cue) had effects as powerful as stereo viewing on task completion times. Moreover, they reported that the effects of head-tracking were *more* powerful than stereoscopic viewing on reducing error rates.

2.3 Shadows as Spatial Cues for Shape and Depth

Yonas [1979] noted the distinction between two types of object shadows: the primary or *attached shadow* that occurs when the shadow of an object is visible on that same object, and the derived or *cast shadow* that occurs when the shadow of one object is visible on a different object. Attached shadows are produced by the different surface orientations of an object that self-shade other surfaces of the same object relative to some light source. In contrast to attached shadows, *cast shadows* are caused by the absence of light created by an occluding object positioned between the light source and a separate, detached, and otherwise illuminated, object surface.

Yonas [1979] cited a number of studies demonstrating that shadows attached to an object can influence the perceived *shape* of that object. For example, the human visual system ‘assumes’ that unseen light sources come from ‘above,’ which is typically true in terrestrial environments. As a result, the objects in Figure 1, apparently illuminated by unseen light sources, are usually perceived as concave ‘dimples’ and convex ‘bumps’ on a surface. That is, the top left object in Figure 1 is seen as a depression receding into the surface whereas the middle top object is seen as a round extrusion above the surface. When the Figure is rotated 180 degrees, the objects ‘reverse’ in shape such that those seen as concave are now seen as convex and vice versa.

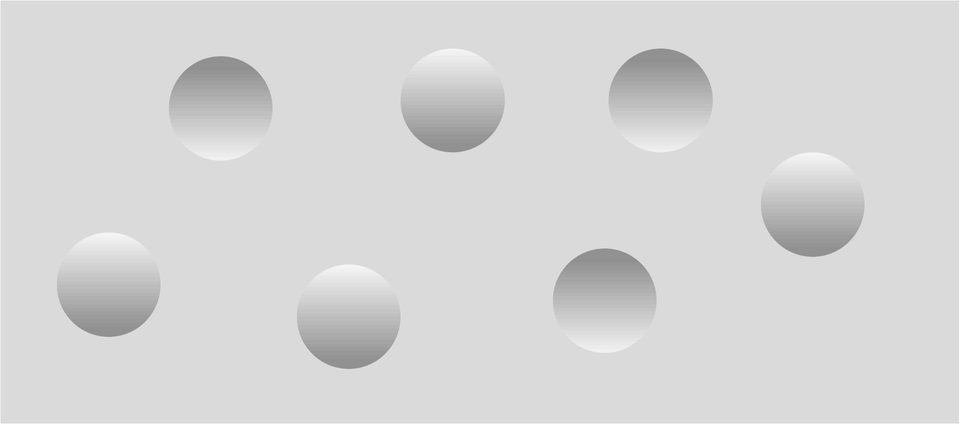

Fig. 1. Attached shadows influence perceived object shape.

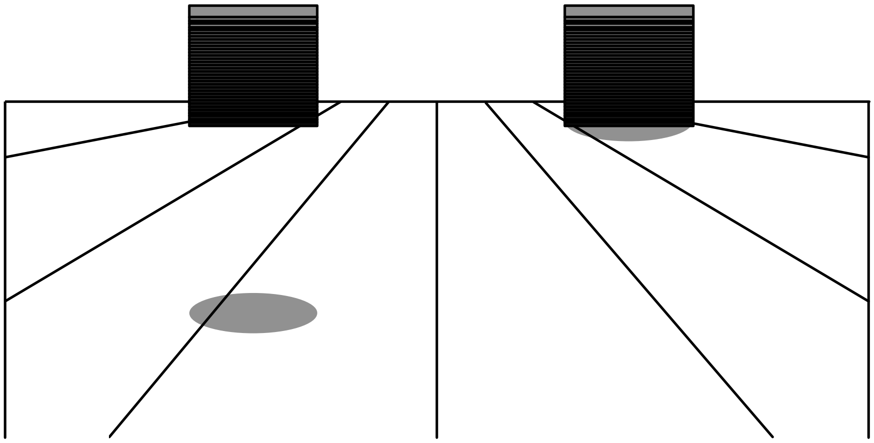

Fig. 2. Cast shadows influence perceived object size, elevation and depth.

It has also been demonstrated that shadows cast by an object can influence the perceived size, elevation, and relative *depth* of that object. In Figure 2 [Yonas et al. 1978], most people ‘see’ the right object as larger in absolute size, and the left object as farther off the ground and closer to them. Yonas' studies demonstrate that attached and cast shadows constitute important sources of information about the shapes, sizes, and layouts and depths of objects in space.

Wanger et al. [1992] manipulated six depth cues while subjects performed object positioning, rotating, and resizing tasks. The manipulated depth cues included: projection (orthographic or perspective); objects casting shadows (on/off); object texture (checkerboard/solid); ground texture (checker/solid); viewpoint motion (on/off); and elevation (object on ground plane/floating in air). Relative to the other five depth cues, the use of shadows had a dominant effect on enhancing the accuracy of performing the object positioning and resizing tasks.

Although Wanger et al. [1992] demonstrated that positioning and sizing objects in 3D space is aided by the presence of shadows cast by the objects, many applications cannot use computer-generated shadows because they are computationally very expensive to render in real-time. Accordingly, the effects of *shadow quality* on the perception of objects in space becomes an important issue. Wanger [1992] conducted three experiments to investigate the effects of *shadow sharpness* (i.e. the accurate rendering of both umbral, or soft shadows, and penumbral regions, or hard shadows) and *shadow shape* (i.e. 'boxy' versus 'true' shadows) on the accuracy of spatial estimation tasks. He concluded that although shadows can be a useful cue to indicate an object's 3D shape, soft shadows can be *detrimental* to determining object shape in the absence of other cues. The implication is that the computationally less expensive 'hard shadowing' computer rendering techniques may be more useful in indicating an object's 3D shape than the computationally more expensive 'soft shadowing' techniques.

Wallach and O'Connell [1953] demonstrated that people can recover 3D form when viewing 2D shadows of moving 3D wire form objects projected onto screens. When the wire forms were stationary, the viewers reported seeing only 2D configurations of lines in the shadow projections. However, when the wire forms were continuously rotated, viewers could accurately recognize the 3D objects represented by the wire forms. Wallach and O'Connell labeled this phenomenon the Kinetic Depth Effect, or KDE. The KDE is well documented as a powerful depth cue and has been demonstrated to exist under relatively impoverished conditions, such as when viewing projections of rotating dot patterns [Braunstein 1976; Todd 1985].

Herndon et al. [1992] demonstrated the capability of shadows to reflect an object's shape by presenting a set of 3D 'shadow widgets' that could be directly manipulated by users to translate, rotate and scale *the objects* that were casting the shadows. The geometries of the shadow widgets were directly related to the geometries of their respective 3D objects. Herndon stated that these shadow widgets reduced the 'cognitive distance' of affecting changes to the corresponding objects because they relied on a concrete visual, real-world metaphor conveying their function.

Pani et al. [1996] conducted experiments investigating whether variations in object orientation affect the ability to imagine the structure and shape of that object's shadow. With basic rectilinear and simple platonic solids, subjects were successful at imagining the shape and structure of shadows when the objects were aligned (i.e. parallel or perpendicular) with the direction of the light source. As the angle became more oblique, performance deteriorated rapidly.

Zhai et al. [1996] described the *partial-occlusion* effect which applies the concept of semitransparent surfaces as a depth cue in 3D target localization tasks. *Occlusion*, or the presence of features hidden behind an object (sometimes referred to as *interposition*), has long been recognized as a powerful depth cue [Schriever 1925; Braunstein et al. 1986; Wickens et al. 1989; Gallimore and Brown 1993; Brown and Gallimore 1995]. However, occlusion is difficult to use with 3D interaction tasks because distal objects may be completely obscured by the more proximal, opaque objects, thereby creating uncertainties about background objects that cannot be viewed. Partial occlusion overcomes this inherent limitation of total occlusion and still provides depth information. Partial occlusion, sometimes referred to as the *silk* effect, is a type of shadowing technique. When a semitransparent surface overlaps another object, the overlapped object is seen in lower contrast (i.e. as if through a silk stocking). Depth information is provided by the 'darkness' of the partial occlusion: the darker the partial occlusion, the more distant the object.

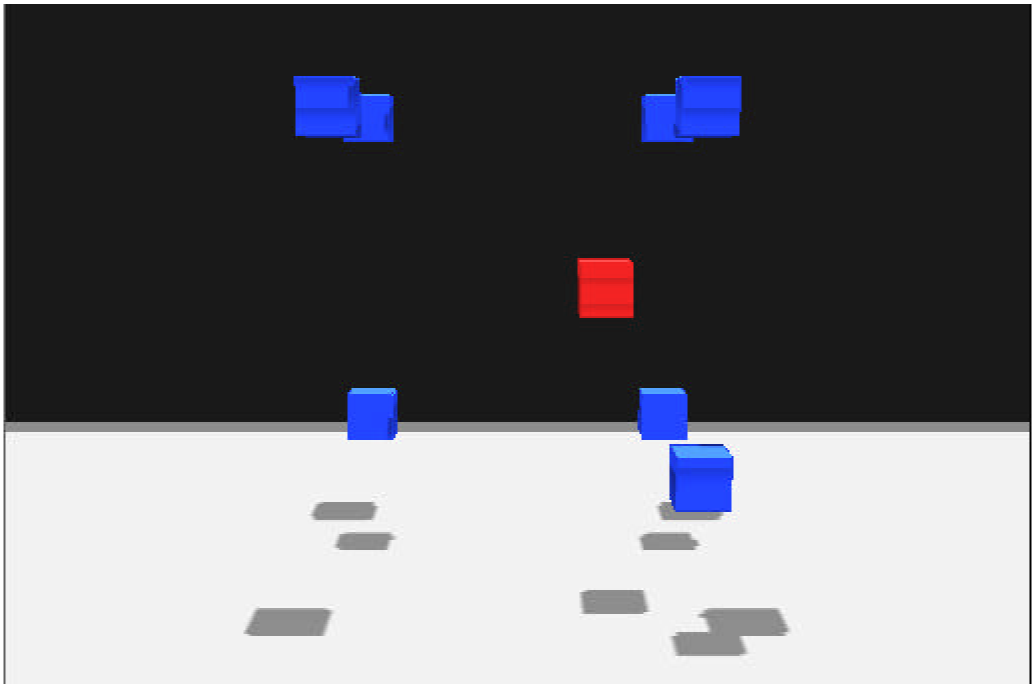

Fig. 3. Example of positioning task trial.

2.4 Experimental Conditions and Hypotheses

There are four experimental conditions in this experiment: *shadows* (on, off); *number of shadow-casting light sources* in the shadows-on condition (one or two); *viewing mode* (stereo, mono); and scene *background* (flat plane, room, 'stair-step' plane). Of primary interest are the relative contributions of the *shadows on* and *stereo viewing* conditions in assisting task performances.

The presence of object shadows has been previously demonstrated to assist positioning and resizing task performances for objects perceived in 3D space (although Wanger et al. [1992] used strictly 2D depth cues). Furthermore, on balance, *cue theory* would suggest that adding the shadows as non-conflicting depth cues should help spatial task performances, although the variant notions of cue theory might counter-indicate the magnitude of the added effect. Accordingly, with respect to both the positioning and resizing tasks, the following hypothesis is proposed: *The presence of objects casting shadows improves task performance*.

The benefits of stereo viewing in performing spatial tasks has previously been demonstrated. Stereopsis is widely accepted as a dominant, physiological enabler for perceiving the location and size of objects in space. In addition, relevant to display design, the *proximity compatibility principle* [Wickens and Carswell, 1995] holds that if a task requires the integration of multiple data variables, they should be bundled in proximity in an integrated display. Positioning (and, arguably, resizing) an object in 3D space based on *x*, *y*, and *z* dimensions perceptually integrated into a stereo display would qualify. Accordingly, we propose the following hypothesis for both the positioning and resizing tasks: *Stereoscopic viewing improves task performance*.

Of secondary interest are performance effects arising from the remaining conditions. We will argue that performance should improve when: (1) the number of shadow-casting light sources increases from one to two; and (2) the background changes from a flat plane to a ‘stair-step’ plane or to a room with walls. Our argument is simple: Both the *weighted additive* and *multiplicative* versions of *cue theory* suggest that adding depth cues serves to assist spatial task performances. Additional light sources cast additional shadows for any given object. We argue that each shadow should provide additional depth information about the corresponding object because the viewer has more information to infer the location (or size) of the object. Thus, task performances should be improved. Similarly, multidimensional background planes also provide additional depth cues about the position and size of objects in space, especially if those objects can be physically manipulated by the viewer. We argue that this holds true for at least two distinct reasons in this experiment. On the one hand, additional depth information is *directly* provided as the objects themselves approach and ‘touch’ the surfaces inclined at right angles. Furthermore, more depth information is *indirectly* provided by the *shadows* approaching (or receding from) the shadowed object along surfaces that intersect at right angles, as opposed to ‘flat’ surfaces. That is, we argue that the ‘corners’ of the room (or stairs) provide more information about the location (and size) of moving objects, either directly by the object’s relative distance, or indirectly, by the object’s shadow, than is true with a flat plane. Thus, we argue that spatial task performances should improve with the room or stair-step background, relative to the plane.

3. METHOD

3.1 Experimental Tasks

3.1.1 *Positioning Task.* Subjects were successively presented with object images, including cubes, spheres, and tetrahedrons, arranged to outline the vertices of symmetrical figures, in the form of cubes or octahedrons. One of the vertex objects was deliberately misplaced in the symmetrical arrangement. The task was to reposition the misplaced vertex object in 3D space (i.e. in the *x*, *y* and *z* dimensions), as quickly and accurately as possible, so as to complete the symmetrical arrangement. Figure 3 depicts a positioning task trial with shadows using cube-shaped vertices arranged in an overall cube figure over a plane illuminated by one light source. In this particular trial, the subject’s task was to move the cube in the center of the display down to the lower front left corner (or vertex) of the overall cube arrangement. Subjects repositioned the vertex objects using a *spaceball*, a six-degrees-of-freedom input device. Subjects manipulated the spaceball in any of the *x*, *y*, or *z* directions, causing the object to ‘fly’ in that corresponding direction.

3.1.2 *Resizing Task.* In the resizing task, subjects were again presented with object images (cubes, spheres, or tetrahedrons) arranged to outline the vertices of a symmetrical figure (a cube or an octahedron) in 3D space. Unlike the positioning task, in the resizing task, the vertex objects were correctly positioned to complete the symmetrical arrangement. However, one of the vertex objects was a different size (either smaller or larger) than the remaining objects. The task was to resize this mismatched object, as quickly and accurately as possible, to make it correspond with the uniform size of the other objects. Figure 4 depicts a resizing task trial using spheres with no cast shadows arranged in an octahedron over stairs.

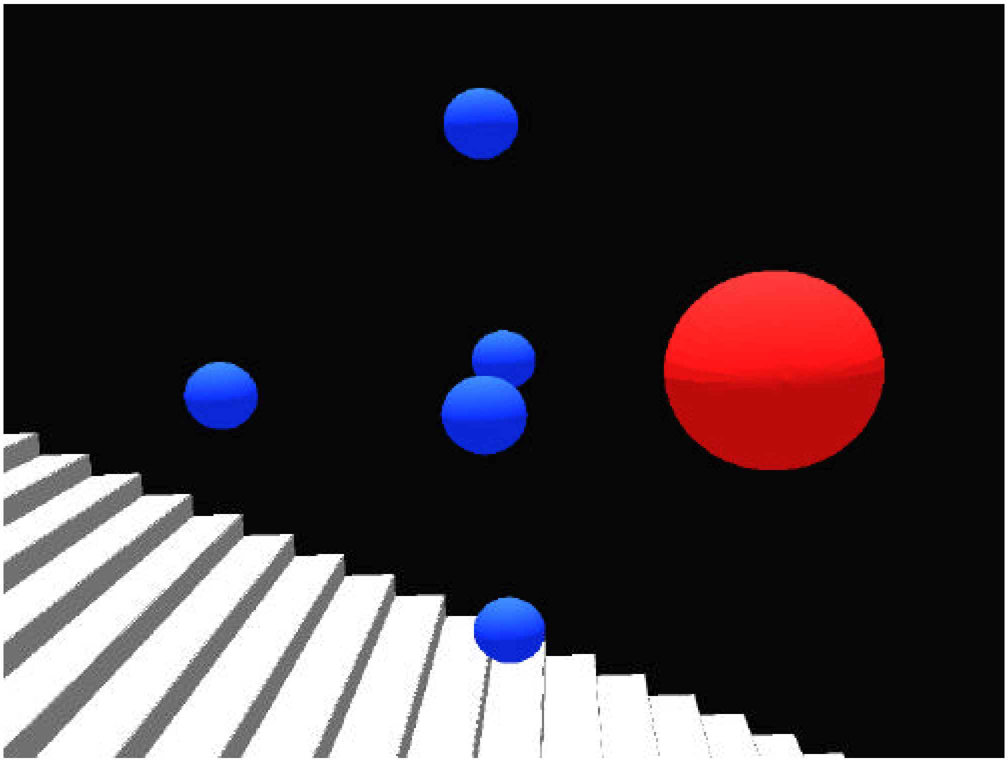

Fig. 4. Example of resizing task trial.

Subjects again used the spaceball input device to resize the objects. Specifically, subjects pushed forward on the spaceball to make the target object smaller, or conversely, they pulled backward on the spaceball to make the object larger. When they perceived that the missized object became the same size as the remaining objects, subjects pressed a button on the spaceball pad, which caused the results of the current trial to be recorded and the next scene to be presented. Subjects used the same button on the spaceball pad to initiate new trials in both the positioning and resizing tasks.

3.2 Experimental Design and Procedure

The experiment used a 2 x 2 x 2 x 3 within-subjects design, manipulating the independent variables: *shadows* (on, off); *number of shadow-casting light sources* (one, two); *viewing mode* (stereo, mono); and scene *background* (flat plane, room, ‘stair-step’ plane). Note that the *number of shadow casting light sources* is a condition that is nested in the *shadows on* condition. There were a total of four sets of 72 trials each, or 288 trials per subject, 144 trials for each task. All subjects viewed the same 288 scenes, although the presentation order varied.

One half of all trials randomly presented shadows on, and the remaining trials presented shadows off. In the shadows on condition, one half of the trials randomly used one light source to produce the shadows, and the remaining trials used two light sources. One half of all trials were viewed in stereo, and the remaining trials were viewed in mono. The stereo and mono conditions were viewed in cohesive 'blocks' of 72 trials each since subjects wore stereoscopic CrystalEyes™ glasses to produce the stereo effect. One third of the background conditions randomly used a flat plane for the scene background; one third used the room; and the remaining third used the 'stair-step' plane. Furthermore, the orientation of the stair-step plane was randomly varied, left-to-right, front-to-back, and right-to-left. Figure 3 presents an example of the flat plane scene background. Examples of the (front-to-back) 'stair-step' plane, and room scene backgrounds are respectively presented as Figures 5 and 6.

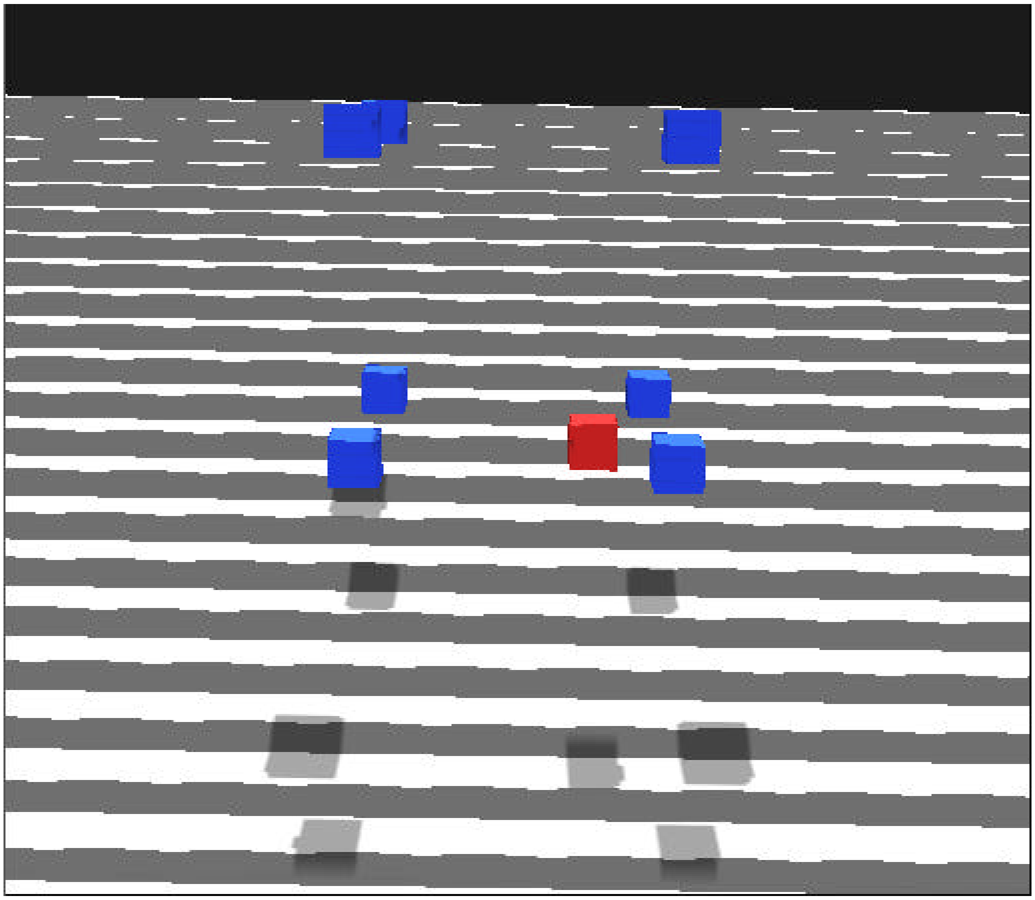

Fig. 5. 'Stair-step' scene background.

In both the positioning and resizing task trials, the location of the misplaced or incorrectly sized vertex was randomly placed with respect to the remaining vertices. In addition, light sources were consistently positioned 'above' the vertex objects and at various angles from the zenith. Moreover, in trials with two light sources, their angle of separation varied from trial to trial. Thus, the relative locations and separation of the light sources changed from trial to trial. Light sources were randomly located and consistently positioned 'from above' to emulate 'real world' conditions in which an observer tries to recover depth information while moving from one (lighted) environment into another (for example, while walking from room to room in a lighted building). Furthermore, we wanted to randomize both the positions of the vertices and the locations of the light sources so that performance effects would be driven by changes in the primary variables of interest.

The dependent variables included *error magnitude* and *response time*. Repeated measures of the dependent variables were automatically recorded by the task trial software. Each subject engaged in 2 sets[6] of 12 practice trials[7] for each of the positioning and resizing tasks for a total of 48 practice trials per subject. Each of the 48 practice trials involved the use of the spaceball input device. The purpose of the practice trials was to expose the subjects to each experimental condition, and to become familiar with the tasks and with the use of the spaceball input device. A small number of subjects requested additional practice trials, specifically to better master the use of the spaceball. We allowed these subjects to re-engage the same practice trials until they were comfortable with the mechanics of the experiment. Data generated from the practice trials were not used in the reported analyses. None of the scenes from the practice trials were repeated in the experimental trials. Following the practice trials, each subject engaged in 2 sets of 72 trials for each of the positioning and resizing tasks, for a total of 144 observations per task per subject. One half of the subjects completed the positioning trials first, and one half completed the resizing task trials first. Subjects were permitted to rest briefly between each set. Subjects completed all positioning and resizing task trial sets in one session that ranged in length from 75 to 120 minutes.

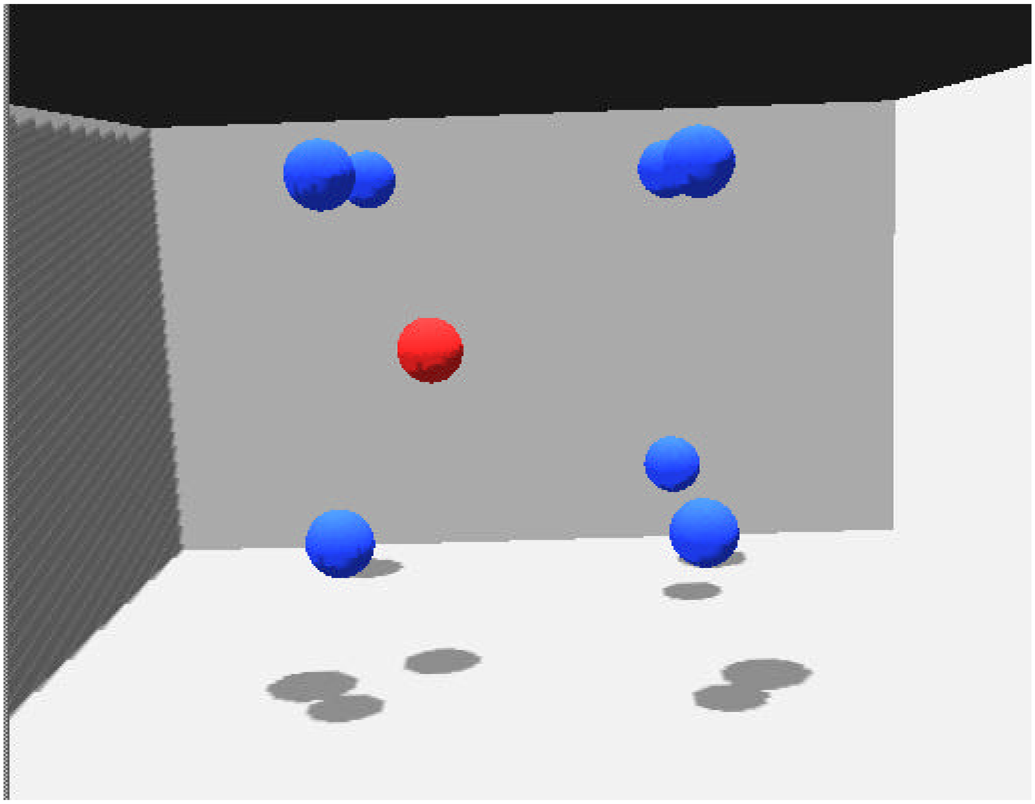

Fig. 6. Room scene background.

---

[6] One set in mono and one set in stereo.

[7] The twelve practice trials represented every possible combination of the lighting and background conditions: shadows on or off x one or two lights (in the shadows on condition) x three background conditions.

The task trial software was developed on a Silicon Graphics, Inc. (SGI) Onyx2 workstation with Infinite Reality2 hardware and IRIX 6.3 system software. The code was written in C and uses OpenGL and GLUT 3.5. Stereoscopic viewing was enabled by wearing 120Hz flicker-free stereoscopic CrystalEyes^{TM} glasses (model no. CE-1), manufactured by StereoGraphics.

As each scene was presented, the target vertex object for positioning or resizing was identified by a reddish color whereas the remaining objects were blue. As soon as the subject manipulated the object with the spaceball, the reddish color changed to match the blue color of the other objects. When finished positioning or resizing the target object, the subject pressed a button on the spaceball pad, which caused the results of the trial to be recorded and the next randomly selected scene in the trial set to be presented.

The volunteer subject population consisted of thirty employees and contractors of the Goddard Space Flight Center. All subjects had professional occupations as engineers, computer programmers or computer scientists. Each subject completed a preliminary questionnaire soliciting demographic information. The mean age of subjects was 40.2 years, with 5.6 mean years of education beyond high school, 15.42 mean years of computer experience, and 18.5 mean years of professional work experience.

3.3 Performance Measures

Two task performance measures were collected: *error magnitudes* and trial completion times (or *response times*). The base unit of measurement for error magnitude was the standard measurement unit in *x, y* and *z* space provided by the software. Specifically, the unit of measure for error magnitude was based on a coordinate system supported by the software for the perspective projection. We arbitrarily allocated exactly 20 (undefined but consistent) units along each of the x, y and z dimensions. That is, the available 3D space (or world in which the objects could move or be resized) was defined as a cube measuring 20 units along each side. Thus, the physical measurement unit (i.e. foot, inch, kilometer) was undefined, but the available ‘distance’ in which to move or resize objects was consistent from scene to scene.

Error magnitude for the positioning task was defined as the Euclidean summation of the three directional errors in the x, y, and z dimensions (i.e. error magnitude $= (e^2_x + e^2_y + e^2_z)^{1/2}$). This metric constitutes the exact distance of the repositioned object from its correct position in three dimensional space. Error magnitude for the resizing task was defined as the absolute value of the error in length of either the radius (in task trials using spheres as vertex objects) or the diagonal (in task trials using cubes or tetrahedrons as vertex objects) of the resized object. Trial completion time was the period of time, measured in milliseconds, from when the scene first appeared until the subject pushed a button on the spaceball, causing the next scene to appear.

3.4 Experimental Results

Table I displays the means and standard deviations of error magnitudes and response times for the two tasks. Note that the main effect of ‘number of lights (one, two)’ is nested under the ‘shadows on’ condition. For the positioning task, the mean overall error magnitude was 1.316 and the mean overall response time was 16.108 seconds. For the resizing task, the mean overall error magnitude was 0.080 and the mean overall response time was 8.985 seconds.

The data was analyzed by fitting a repeated measures multivariate analysis of variance model (MANOVA) to the experimental observations. The MANOVA model tested each of the four main effects (shadows on/off, viewing mode, number of shadow-casting lights, and background) on the error magnitude and response time dependent variables. The data from the two tasks were segregated and analyzed separately. Type III sums of squares are reported for main effects that were also components of significant interaction effects. Type III sums of squares take into account the influence of main effects before the influence of interaction effects are computed. Thus, a significant interaction effect does not nullify the statistical significance of any corresponding main effect.

Table I. Means and standard deviations of error magnitudes and response times.

	Positioning Task				Resizing Task			
	Error Magnitude		Response Time (secs.)		Error Magnitude		Response Time (secs.)	
Use of Shadows:	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev
On	1.225	1.399	16.634	12.025	0.079	0.099	9.198	5.747
Off	1.406	1.618	15.581	11.202	0.081	0.096	8.773	5.561
Viewing Mode:	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev
Stereo	0.737	0.632	14.125	9.838	0.073	0.090	8.559	5.637
Mono	1.894	1.877	18.091	12.883	0.088	0.103	9.411	5.648
Number of Lights:	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev
One	1.129	1.291	15.623	10.490	0.076	0.085	8.924	5.373
Two	1.316	1.489	17.590	13.249	0.082	0.110	9.458	6.072
Scene background:	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev
Plane	1.279	1.408	15.783	10.913	0.087	0.092	8.817	5.557
Room	1.225	1.410	15.660	10.803	0.068	0.083	9.048	5.765
Stairs	1.443	1.700	16.880	13.014	0.086	0.113	9.091	5.651

3.4.1 *Positioning Task Results*. In the positioning task performance data, there were significant differences (at the 95% confidence level) in the mean values of the dependent variables (error magnitude and response time) as a function of all four main effects: use of shadows; viewing mode; number of shadowing lights; and scene background. In addition to these significant main effects, there was also a significant interaction between the scene background and the number of shadow-casting light sources.

The use of shadows significantly affected subjects' object positioning performances in the omnibus MANOVA model (F (2, 4280) = 19.79; p = 0.0001). With shadows on, subjects were more accurate (the 'shadows on' condition mean error magnitude is 1.225; F (1, 4281) = 19.74; p = 0.0001) than they were without the shadows (the 'shadows off' condition mean error magnitude is 1.406). However, subjects *took longer* positioning the vertex objects using shadows (the mean 'shadows on' positioning response time is 16.634 seconds; F (1, 4281) = 13.59; p = 0.0002) than did subjects without the shadows (the mean 'shadows off' condition response time is 15.581 seconds).

The viewing mode also significantly impacted subjects' positioning task performances in the omnibus MANOVA model (F (2, 4280) = 442.28; p = 0.0001). Subjects viewing the vertex objects in stereo were more accurate (the mean stereo viewing condition error magnitude is 0.737; F (1, 4281) = 804.56; p = 0.0001) than were subjects viewing the vertex objects in mono (the mean mono viewing condition error magnitude is 1.894). Furthermore, subjects viewing the objects in stereo were faster performing the positioning task (mean stereo viewing condition response time is 14.125 seconds; F (1, 4281) = 193.04; p = 0.0001) than were subjects viewing the objects in mono (the mean mono condition response time is 18.091 seconds).

The number of shadow-casting lights also significantly impacted subjects' positioning task performances in the omnibus MANOVA model (F (2, 4280) = 11.50; p = 0.0001). Subjects viewing the vertex objects shadowed by two lights sources were *less* accurate (the mean two lights condition error magnitude is 1.316; F (1, 4281) = 6.24; p = 0.0125) than were subjects viewing the vertex objects shadowed by one light source (the mean one light condition error magnitude is 1.129). Furthermore, subjects viewing the objects shadowed by two light sources took *more* time performing the positioning task (mean two light condition response time is 17.59 seconds; F (1, 4281) = 19.73; p = 0.0001) than did subjects viewing the objects shadowed by one light source (the mean one light condition response time is 15.623 seconds).

The scene background also significantly impacted subjects' positioning task performances in the omnibus MANOVA model (F (4, 8560) = 8.99; p = 0.0001). Subjects viewing the vertex objects in a room (the mean room scene background condition error magnitude is 1.225; F (2, 4281) = 12.41; p = 0.0001) or over a plane (the mean plane scene background condition error magnitude is 1.279) were more accurate than were subjects viewing the vertex objects over stairs (the mean stairs condition error magnitude is 1.443). The error magnitude difference between the room and plane conditions was not significant. Furthermore, subjects viewing the objects over stairs (the mean stairs scene background condition response time is 16.88 seconds; F (2, 4281) = 8.44; p = 0.0002) took more time performing the positioning task than did subjects viewing the objects in a room (the mean room scene background condition response time is 15.66 seconds) or over a plane (the mean plane condition response time is 15.783 seconds). The difference in the response times between the room and plane scene background conditions was not significant.

In addition to the main positioning task effects, there was also a statistically significant interaction between the scene background and the number of shadow-casting light sources in the omnibus MANOVA model (F (8, 8560) = 2.78; p = 0.0045). This interaction effect was significant with respect to error magnitude (F (4, 4281) = 4.51; p = 0.0012), but not with respect to response time (F (4, 4281) = 1.55; p = 0.1847). Figure 7 graphically presents this interaction effect.

3.4.2 *Resizing Task Results*. In the resizing task performance data, there were significant differences (at the 95% confidence level) in the mean values of the dependent variables (error magnitude and response time) as a function of all four main effects: use of shadows; viewing mode; number of shadow-casting light sources; and scene background. In addition to these significant main effects, there was also a significant interaction between the viewing mode and the number of shadow-casting light sources.

The use of shadows significantly affected subjects' object resizing performances in the omnibus MANOVA model (F (2, 4282) = 6.61; p = 0.0014). There was no difference in accuracy (the 'shadows on' condition mean error magnitude is 0.079; F (1, 4283) = 0.71; p = 0.3982; whereas the 'shadows off' condition mean error magnitude is 0.081). However, subjects *took longer* resizing the vertex objects using shadows (the mean 'shadows on' resizing response time is 9.918 seconds; F (1, 4283) = 12.08; p = 0.0005) than did subjects without the shadows (the mean 'shadows off' condition response time is 8.773 seconds).

The viewing mode also significantly impacted subjects' resizing task performances in the omnibus MANOVA model (F (2, 4282) = 24.52; p = 0.0001). Subjects viewing the vertex objects in stereo were more accurate (the mean stereo viewing condition error magnitude is 0.073; F (1, 4283) = 18.59; p = 0.0001) than were subjects viewing the vertex objects in mono (the mean mono viewing condition error magnitude is 0.088). Furthermore, subjects viewing the objects in stereo were faster performing the resizing task (mean stereo viewing condition response time is 8.559 seconds; F (1, 4283) = 33.50; p = 0.0001) than were subjects viewing the objects in mono (the mean mono condition response time is 9.411 seconds).

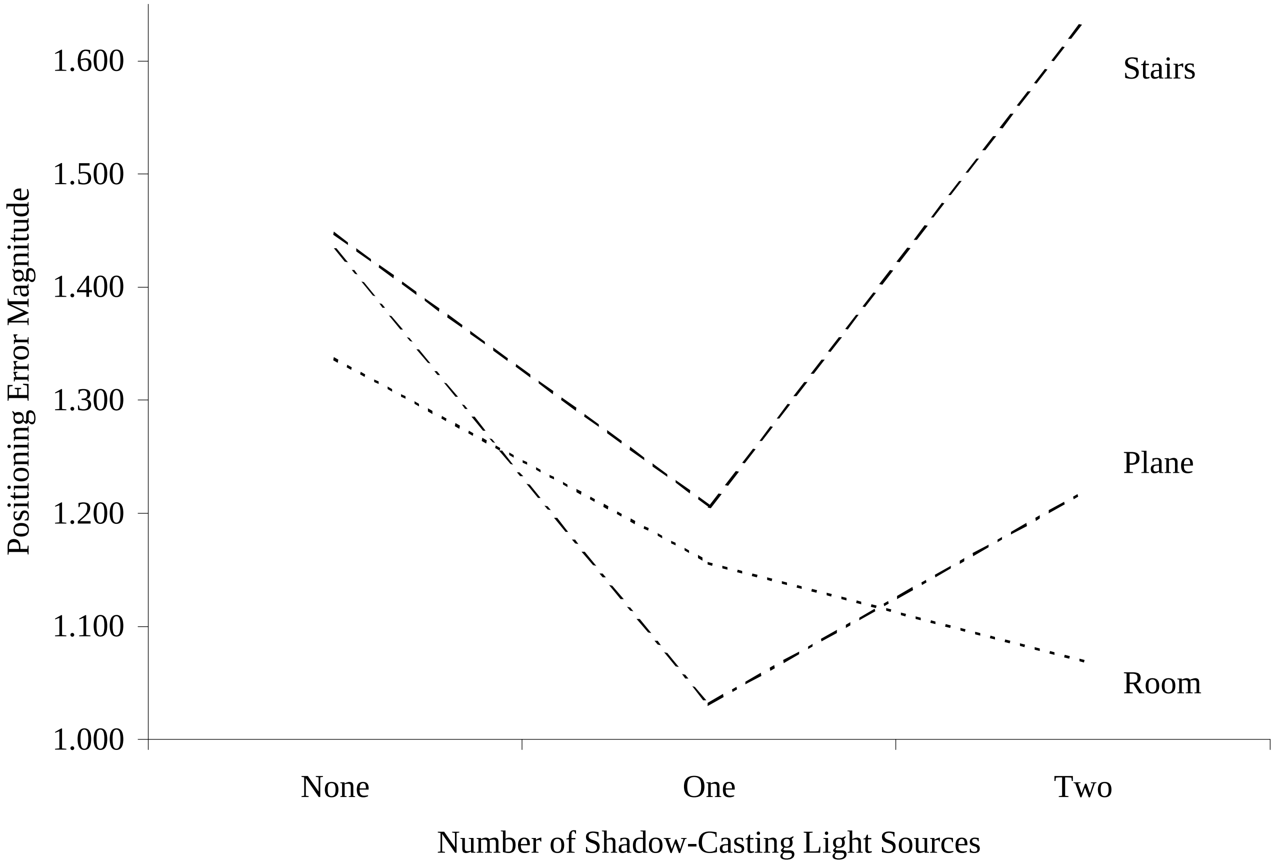

Fig. 7. Interaction of scene background with the number of shadow-casting lights on positioning accuracy.

The number of shadow-casting lights also significantly impacted subjects' resizing task performances in the omnibus MANOVA model (F (2, 4282) = 5.17; p = 0.0057). There was no difference in accuracy with shadows created by two lights (the mean two lights condition error magnitude is 0.082; F (1, 4283) = 1.79; p = 0.1805) compared to one light source (the mean one light condition error magnitude is 0.076). However, subjects viewing the objects shadowed by two light sources took more time performing the resizing task (mean two light condition response time is 9.458 seconds; F (1, 4283) = 9.03; p = 0.0027) than did subjects viewing the objects shadowed by one light source (the mean one light condition response time is 8.924 seconds).

The scene background also significantly impacted subjects' resizing task performances in the omnibus MANOVA model (F (4, 8564) = 9.96; p = 0.0001), but only with respect to resizing accuracy, not response time. Subjects viewing the vertex objects in a room (the mean room scene background condition error magnitude is 0.068; F (2, 4283) = 17.72; p = 0.0001) were more accurate than subjects in either the plane condition (the mean plane scene background condition error magnitude is 0.087) or the stairs condition (the mean stairs condition error magnitude is 0.086). The differences in the response times among the three scene background conditions were not significant (F (2, 4283) = 1.81; p = 0.1646). The mean response time resizing objects over a plane was 8.817 seconds, over a room was 9.048 seconds, and over stairs was 9.091 seconds.

In the resizing task data, there was a significant interaction effect between viewing mode and the number of shadow-casting lights (F (4, 8564) = 3.201; p = 0.0123) with respect to both accuracy (F (2, 4283) = 3.22; p = 0.0402) and response time (F (2, 4283) = 3.58; p = 0.0281). Figures 8 and 9 graphically present these interactions.

4. DISCUSSION

The findings of this study are mixed with respect to the first hypothesis that: "the presence of objects casting shadows improves performance in spatial tasks." The presence of object shadows did not uniformly improve performance in both spatial tasks. In the positioning task, accuracy was improved, but response times were longer. This is symptomatic of speed-accuracy performance tradeoffs. It is difficult to discern whether the shadows had a direct beneficial influence on object positioning, or whether this influence was indirect because the subjects spent more time on the task. In the resizing task, the shadows had no effect on accuracy, and, again, response times were longer.

In discussing these findings, we should note that this group of subjects likely has heightened spatial abilities relative to the general population, although we did not capture this metric in any formal fashion. However, they were all working professionals within a premier scientific governmental organization, many with advanced degrees. Whether these findings generalize to the population at large is debatable. At any rate, one could argue that any performance effects fostered by the spatial cues would be *more* pronounced with this group than within the general population. That is to say, if there is no observable effect with these subjects, there would likely be no effect within the general population. On the other hand, the presence of an effect with these subjects would not necessarily carry over to the population at large.

Furthermore, the random locations of the light sources probably made the task more difficult than had they been fixed from trial to trial. Light source positions were deliberately randomized to eliminate the specific location as a factor affecting subjects' performances. Furthermore, there was the rationale to emulate 'real world' conditions, as, for example, when individuals navigate through various environments with light sources placed at different angles and positions. In hindsight, on balance, it likely would have been better to fixate the lighting positions. We suggest that this is a possible design flaw in the experiment [Schneider and Shiffrin 1977; Shiffrin and Schneider 1977] and we encourage future researchers in this area to carefully consider the lighting positions as a deliberate factor in their research.

In addition, we should note that although the size of the effects are small relative to the variability, the tight, repeated measures design produced strong main effects in the omnibus

MANOVA models, with $p < 0.0001$ in all cases in the positioning task and $p < 0.006$ in all cases in the resizing task. Each of the thirty subjects engaged in 144 repeated measure trials for each task, with the trial scenes presented in varied orders. We deliberately chose this tightly-coupled design to evoke the most powerful effects from a relatively small number of subjects. Thus, it is reasonable that user interface designers might expect similar results, given users with similar characteristics engaged in similar tasks.

The presence of the object shadows did improve subjects' accuracy in positioning objects, but at the expense of taking more time to finish the trials. In spite of longer response times with the shadows, one could argue that the *quality* of their positioning task performance was enhanced since they were more accurate. However, they took *longer* to position the objects using the shadows, even though they were instructed to work as accurately *and* quickly as possible. Indeed, in informal conversations with the subjects immediately following the positioning trials, many stated that they only used the additional shadowing cue *after* they had used other perceptual cues to place the mispositioned object in the perceived correct location. That is, they used the shadowing cue as a 'final check' to more precisely calibrate the position of the misplaced vertex object. However, it is relevant to note, that, in these conversations, many subjects expressed great enthusiasm for the utility of the shadows in assisting positioning task performance. Considering the experimental data in conjunction with this anecdotal evidence suggests that many subjects used the shadow cues to more precisely 'fine tune' an initial positioning decision.

However, in the resizing task, the presence of the object shadows did not improve performance accuracy at all, and, they again *took longer* to complete the trials. This degraded performance in resizing the objects in the presence of cast shadows belies the fact that many subjects expressed confidence that the shadows aided their performance. In fact, as a group, they performed better with no shadows at all. Evidently, they again attempted to use the shadowing cue for additional final assistance in calibrating the object's size, but it did not significantly improve their accuracy (even though many thought that it did).

Due to the improved accuracy/increased response time tradeoffs, it is problematic to assert that the presence of the shadows assisted positioning performance (but not resizing performance) in general. But why did positioning accuracy improve with no commensurate gain in resizing accuracy? We speculate that positioning accuracy was enhanced due to the nature of that task. Specifically, the positioning task afforded the subjects a larger degree of control over the *motion* of the objects, which has been demonstrated to be a powerful depth cue in its own right [Hubona et al. 1997]. We suggest that the combination of 'flying' the objects around the virtual space, in conjunction with the depth information provided by the moving shadows, provided a richer set of depth information than did the shadows cast by the stationary resized objects.

In contrast to the mixed results in subjects' performances tied to the shadowing cue, stereo viewing, compared to mono viewing, uniformly improved all measures of performance for both spatial tasks, as speculated by the hypotheses. Subjects were both faster and more accurate positioning and resizing the objects in stereo. Furthermore, when subjects were asked whether they thought the stereo viewing mode was helpful, they typically responded in the affirmative. Responses such as: "Sure!" and "Of course!" were common. Clearly, the empirical evidence, as well as the subjects' verbal responses, indicate that stereo viewing is a powerful cue that had a dominant influence on the spatial task performances, relative to the shadowing cue.

However, perhaps the data with respect to the number of shadow-casting light sources is more revealing regarding the utility of the shadows in assisting these task performances. The hypotheses speculate that both positioning and resizing task performances would improve as the number of shadow-casting light sources increases from one to two. This was not the case. Positioning accuracy significantly improved as the number of shadow-casting light sources increased from none (mean error magnitude = 1.406) to one (mean error magnitude = 1.129), but then significantly declined from one light source to two (mean error magnitude = 1.316). In fact, positioning accuracy with two shadow-casting light sources was not significantly different from that with no shadows at all. Positioning response times were not significantly different with no shadows (mean response time = 15.58 seconds) compared to shadows produced by one light (mean response time = 15.623 seconds), but when objects were shadowed by two lights (mean response time = 17.59 seconds), subjects were significantly *slower* in positioning the objects than when the objects were shadowed with one light or when there were no shadows at all.

Resizing accuracy was not significantly different with either no shadows (mean error magnitude = 0.081), shadows from one light source (mean error magnitude = 0.076), or from two (mean error magnitude = 0.082) light sources. Similarly, resizing response times were not significantly different with no shadows (mean response time = 8.773 seconds) or with shadows from one light source (mean response time = 8.924 seconds), although subjects were significantly slower when objects were shadowed by two light sources (mean response time = 9.458 seconds) than with either one light or no shadows at all.

These results indicate that, contrary to our speculation, more shadow-casting light sources did not provide better and more comprehensive spatial information about the objects' positions and sizes. Our *a priori* assumption that additional shadowing light sources would provide linearly increasing amounts of spatial information about the corresponding objects was flawed. In fact, objects shadowed by two light sources often fostered *ambiguous* and *conflicting* spatial information about the objects. Recall that the position of the light sources varied from trial to trial. Although subjects reported they could more effectively infer the position of an object shadowed by one light source than with no shadows, a number of subjects expressed confusion about the positions of both the lights and the objects when they were shadowed by two light sources. In these cases, the shadows did not help them infer the relative positions and sizes of the corresponding objects. In ensuing discussions, subjects speculated that two shadow-casting light sources might have helped more had they been uniformly fixed in their angular and distance orientations in each scene.

With the scene background conditions, our results again deviated from those hypothesized, perhaps because of the same erroneous assumption of linearly increasing amounts and quality of spatial information as the scene background becomes more 'multidimensional' (i.e. changes from a plane to stairs or to a room). In the positioning task, the stair-step scene background was problematic for performance, both in terms of accuracy and response time. There was no difference in positioning accuracy, nor response time, for the plane compared to the room. In the resizing task, accuracy was improved with a room scene background, compared to either the plane or stairs. There were no significant resizing response time differences among either of the three scene background conditions.

The presence of two shadows and the use of more complex stair-step background scenes were clearly problematic for both task performances. Wickens [1992a] and Wickens and Carswell [1995] discuss four forces that can help or hinder improved performances in tasks requiring the integration of proximal depth cues. Specifically, they discuss the increased perceptual *information access cost* (IAC) that can *disrupt* performance [p.484]: "when images overlap and cannot be easily discriminated by source differences" such that "it becomes difficult to perceptually parse, or segregate, the objects from one another." They state that this "*clutter*" can "disrupt the movement of visual attention to the indicator, often imposing greater uncertainty as to where that target is located." In hindsight, we feel that the overlapping depth cues fostered by multiple shadows generated by unseen light sources, and especially in conjunction with the more fragmented stair-step scene backgrounds, served to create such "*disruptive clutter*" in the intended depth cue integration tasks.

Notice the significant interaction (see Figure 7) between the scene background and the number of shadow-casting light sources in positioning task accuracy. Introducing one shadow-casting light improves positioning accuracy, regardless of scene background. Recall that although the shadows (on or off) condition had no effect on resizing performance, the number of light sources did have a significant effect. However, introducing the second shadow-casting light *impairs* performance, greatly with the stairs, but also over a plane. However, over the room, the second light serves to *improve* positioning accuracy.

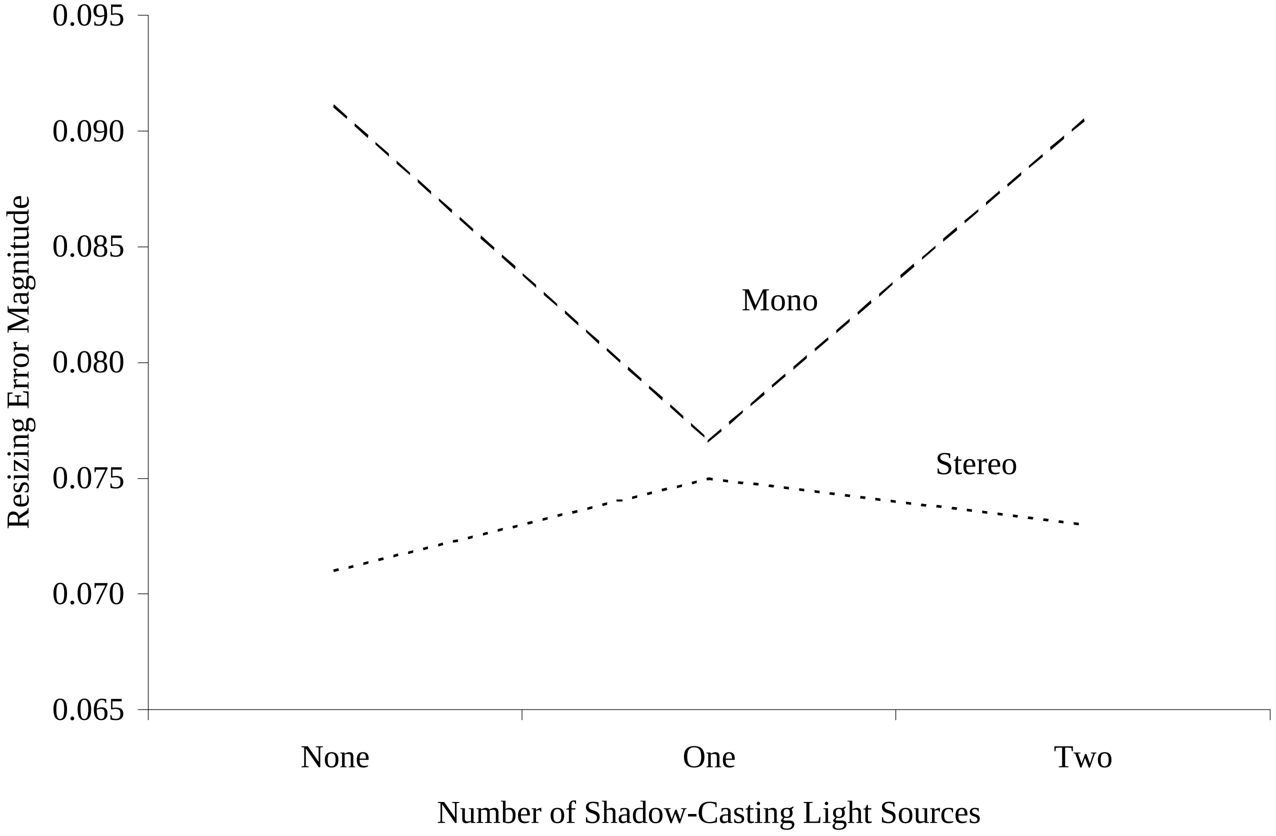

Fig. 8. Interaction of viewing mode with the number of shadow-casting lights on resizing accuracy.

We argue that positioning the objects in a room does achieve a *dimensional integrality* effect as described by Wickens et al. [1994] such that the task of accurately positioning objects is enhanced. Specifically, the bounding walls of the room provide additional dimensional integration depth cues, regardless of whether the scene is displayed in mono or stereo, or whether shadows are present or absent. As the objects are moved within the room, there is less ambiguity regarding the 3D positioning of the objects, even in the presence of two shadow-casting light sources. We suggest there is less ambiguity for two reasons. As the objects approach a 'wall' in the room, the shadow associated with that object is immediately apparent as the shadow 'climbs' up or down the wall to 'touch,' or converge with, the object. Thus, the object's relative 3D position is very apparent even if the object has an additional shadow on the floor or elsewhere. Furthermore, as the object itself 'touches' and 'passes into'[8] the wall (with or without shadows), the resulting occlusion of part of the object provides an additional powerful depth cue that is not evident by moving objects over a plane or over stairs. Thus, we suggest that bounding the object *inside* a room, with walls in 3D space, provides additional powerful dimensional integrality depth cues to locate that object's position relative to other objects in the room.

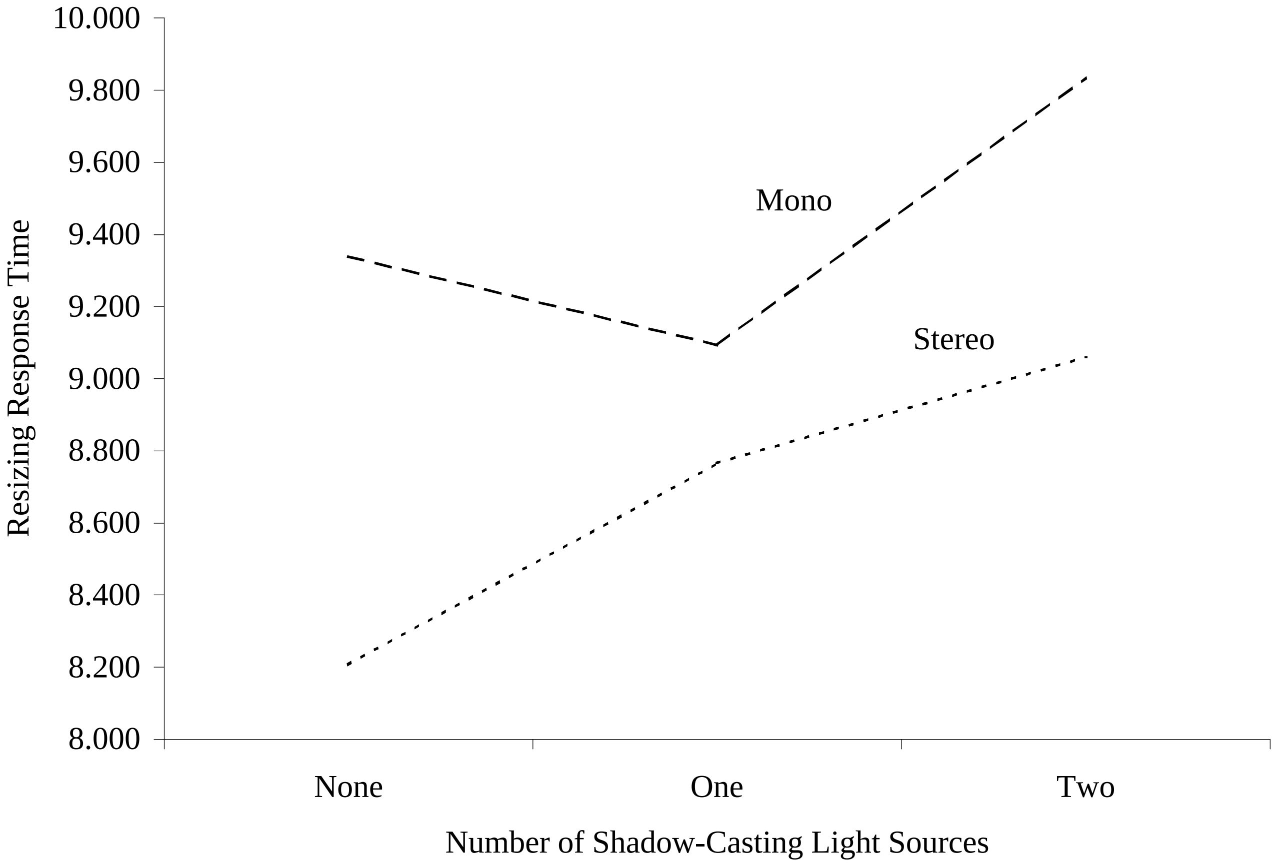

Fig. 9. Interaction of viewing mode with the number of shadow-casting lights on resizing response time.

---

[8] Object collisions were not modeled, thus objects could 'pass through' walls and other objects.

5. CONCLUSIONS

This evidence clearly suggests that stereo viewing is more powerful than the use of object shadows in providing depth cues about the relative size and position of objects in space. But how does this evidence relate to alternate versions of *cue theory*? That is, what does this data indicate about how visual depth cues are ‘combined’ from a human information processing perspective so as to provide an integrated cognitive impression regarding some spatial characteristic of the perceived object, such as relative or absolute size, distance, elevation, and so forth?

Figures 8 and 9, which depict the interaction of viewing mode with the number of shadow-casting light sources, provide some clues in this regard. In Figure 8, notice that the introduction of a single shadow-casting light source precipitates improved resizing accuracy in the mono viewing mode, but does not improve accuracy in the stereo mode. Introducing the second light source impairs resizing accuracy in mono, and does not significantly change the level of resizing accuracy in stereo. In Figure 9, a similar pattern emerges with respect to resizing response time. Introducing one shadow-casting light, and then two, causes resizing response time to successively increase in the stereo mode. In the mono viewing mode, introducing one shadow-casting light reduces the response time, while adding the second light causes response time to increase.

Because viewing mode, and object shadows produced by the light sources, are the dominant depth cues in this experiment, we can draw some inferences regarding alternate versions of cue theory. Based on the empirical data in this experiment, stereo viewing is dominant over the number of shadows. Combining stereo viewing with one, and then two shadows, causes both resizing accuracy and response time to *degrade*. That is, task performance while viewing the objects in stereo *gets worse* as shadow-casting lights are added. This observation clearly does not support either the *additive* [Bruno and Cutting 1988] or *multiplicative* [Sollenberger 1993] versions of cue theory. These theories predict improved performance, either additively, or multiplicatively, as more depth cues are added to the scene. The data is more consistent with the *vetoing* and *strong fusion* mechanisms [Johnston et al. 1993]. The *vetoing* mechanism suggests that the more dominant depth cue (i.e. stereo) simply overrides the weaker cue (i.e. shadows), such that the combined effect is no stronger than that produced by the more powerful cue alone. The *strong fusion* model suggests that depth information is processed separately by each cue, and then combined in a non-linear fashion, that is contingent on the situation and the task. This perspective also helps to explain the differential effects of introducing shadows in the mono and stereo viewing conditions on task performance.

What about the practical implications of these findings? Overall, the findings indicate that spatial task performances may be hindered through the use of multiple cast shadows in complex environments. This may be especially true for tasks involving the *positioning* of objects in space. The implication for user interface design is clear: Whereas the use of singular shadows in less complex scenes may provide useful depth information so as to assist positioning task performances, these performance benefits rapidly degrade as the number of shadows, or scene complexity, increases. Additional findings relevant to the design of 3D user interfaces include:

- The addition of one light source over no lighting enhanced object positioning accuracy regardless of background scene configuration;
- Adding a second light source never enhanced positioning performance, but often impaired that performance;

- Introducing one light source to a monoscopically viewed scene enhanced resizing accuracy to a level near that of stereoscopically viewed resizing;
- Adding a second light source rendered monoscopic resizing performance as impoverished as no lighting;
- Stereoscopic viewing was a dominant depth cue; stereoscopic viewing was superior to monoscopic viewing, and to any shadow condition, for enhancing positioning and resizing accuracy and response times; and
- The Johnston et al. [1993] research on vetoing and strong fusion deserve further examination as these results do not support the additive or multiplicative versions of cue theory.

5.1 Considerations for Future Research

More research is needed to assess the effects of object shadows in assisting spatial task performances. There has been little research in this area, perhaps due to practical constraints imposed by the lack of suitably powerful hardware. This experiment represents one of the ‘first looks’ at this topic. To the extent that future researchers build upon this work, we suggest two modifications to the experimental design.

One suggestion is that future experiments use fixed lighting positions. As previously noted, varying the positions of the lights probably mitigated the helpful effects of the object shadows on task performances. Previous research demonstrates that fixed lighting positions are optimally helpful to tasks of this nature [Kersten et al. 1996, 1997]. This is because the observer infers information about object position by relating the position of the shadows to the position of the light source(s) relative to the object. When the actual light sources are unseen, the assumption of stationary lighting positions fosters a more consistent geometry that assists the ability to recover information about object location. Thus, one could argue that moving the lights represents the ‘weak case’ to assess the effect of object shadows, whereas fixating the lights represents the ‘strong case.’ Since the results of this ‘weak case’ experiment were mixed with respect to the impact of the shadows on the accuracy and speed of object positioning and resizing, future research could utilize the ‘strong case’ of fixed lighting positions to look at performance effects at the other end of the spectrum.

Another suggested modification to the current study relates to the argument that scaling, or resizing, objects in space is not obviously a task involving depth perception, especially when the target and referent objects are located at equal distances from the observer.[9] In this study, the target and referent objects were not consistently located at different distances from the observer. The experimental design might be stronger if all trials were deliberately structured such that target and referent objects were consistently located at disparate distances. This scenario would more directly require the use of depth perception, in conjunction with the use of cognitive size constancy mechanisms, to compare the perceived relative sizes of the objects.

One should note that most existing research on the influence of shadows on perception focuses on the perception of *shape*, rather than on the perception of the relative *depth* of objects in space. More research is needed to achieve a detailed understanding of precisely what depth information is made available by shadows and how an observer may recover this information. In addition, we agree with Wanger et al. [1992] that there is the need for more research to investigate and develop a comprehensive taxonomy of: (1) spatial manipulation tasks; (2) types

---

[9] However, the findings that stereo viewing uniformly assisted both the accuracy and speed of object resizing would suggest that the resizing task was influenced by depth perception mechanisms.

(i.e. dimensions) of spatial information; and (3) the visual cues that provide this information, such that more effective 3D applications and accompanying user interfaces might be developed. With the demonstrated power of stereo viewing as a dominant cue, more research investigating the efficacy of a wider range of pictorial cues against this baseline cue is warranted. Furthermore, with the advent of the world wide web and the widespread adoption and use of interactive, visual, animation-enabling languages, such as VRML and java, more research is needed to determine how best to display 3D pictures and scenes on the web such that more effective use may be made of globally-available internet technology.

REFERENCES

ANDRE, A.D., WICKENS, C.D., MOORMAN, L. AND BOSCHELLI, M. 1990. *SID International Symposium Digest of Technical Papers. 21*, 347-350.

ARTHUR, K., BOOTH, K., AND WARE, C. 1993. Evaluating 3D task performance for fish tank virtual worlds. *ACM Transactions on Information Systems. 11*, 3 (July), 239-265.

BRAUNSTEIN, M.L. 1976. *Depth perception through motion*. New York: Academic.

BRAUNSTEIN, M.L., ANDERSON, G.J., ROUSE, M.W., AND Tittle, J.S. 1986. Recovering viewer-centered depth from disparity, occlusion and velocity gradients. *Perception Psychophys. 40,* 216-224.

BRIDGES, A.L. AND REISING, J.M. 1987. Three-dimensional stereographic pictorial visual interfaces and display systems in flight simulation. *SPIE Proceedings. 761*, 102-109.

BROWN, M.E. AND GALLIMORE, J.J. 1995. Visualization of three-dimensional structure during computer-aided design. *International Journal of Human-Computer Interaction. 7*, 37-56.

BRUNO, N. AND CUTTING, J.E. 1988. Minimodularity and the perception of layout. *Journal of Experimental Psychology: General. 117*, 161-170.

BÜLTHOFF, H.H. AND MALLOT, H.A. 1988. Integration of depth modules: Stereo and shading. *Journal of the Optical Society of America, A. 5*, 10, 1749-1758.

CLARK, J.J. AND YUILLE, A.L. 1990. *Data Fusion for Sensory Information Processing Systems*. Kluwer Academic Publishers, London.

CUTTING, J.E., BRUNO, N., BRADY, N.P. AND MOORE, C. 1992. Selectivity, scope, and simplicity of models: A lesson from fitting judgments of perceived depth. *Journal of Experimental Psychology: General*, *121*, 364-381.

FOARD, C.F. AND NELSON, D.C.K. 1984. Holistic and analytic modes of processing: the multiple determinants of perceptual analysis. *Journal of Experimental Psychology: General*, *113*, 94-111.

GALLIMORE, J.J. AND BROWN, M.E. 1993. Visualization of 3-D computer-aided design objects. *International Journal of Human-Computer Interaction*. *5*, 361-382.

GARNER, W.R. AND FELFOLDY, G.L. 1970. Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology, 1,* 225-241.

HERNDON, K.P., ZELENIK, R.C., ROBBINS, D.C., CONNER, D.B., SNIBBE, S.S., AND VAN DAM, A. 1992. Interactive shadows. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. ACM, New York 1-6.

HUBONA, G.S., SHIRAH, G.W., AND FOUT, D.G. 1997. The effects of motion and stereopsis on 3D visualization. *International Journal of Human-Computer Studies, 47*, 5, 609-627.

JOHNSTON, E.B., CUMMING, B.G., AND PARKER, A.J. 1993. Integration of depth modules: Stereopsis and texture. *Vision Research. 33,* 5/6, 813-826.

KELSEY, C.A. 1993. Detection of visual information. In *The Perception of Visual Information*, W.R. HENDEE AND P. WELLS, Eds. New York: Springer-Verlag, 30-51.

KERSTEN, D., MAMASSIAN, P., KNILL, D.C., AND BÜLTHOFF, I. 1996. Illusory motion from shadows. 1996. *Nature, 379,* 31.

KERSTEN, D., MAMASSIAN, P., AND KNILL, D.C. 1997. Moving cast shadows induce apparent motion in depth. *Perception, 26,* 171-192.

LIU, Y. AND WICKENS, C.D. 1992. Use of computer graphics and cluster analysis in aiding relational judgment. *Human Factors*. *34,* 165-178.

LOCKHEAD, G.R. 1972. Processing dimensional stimuli: a note. *Psychological Review*, *79*, 410-419.

MCALLISTER, D.F., Ed. 1993. *Stereo Computer Graphics and Other True 3D Technologies*. Princeton University Press, Princeton, N.J.

MCWHORTER, S.W., HODGES, L.F., AND RODRIGUEZ, W.E. 1991. *Comparison of 3D display formats for CAD applications* (Tech. Report GIT-GVU-91-04). Atlanta: Georgia Institute of Technology, Graphics, Visualization and Usability Center.

OVERBEEKE, C.J. AND STRATMANN, M.H. 1988. Space through movement. Ph.D. thesis, Delft Univ. of Technology, Delft, Netherlands.

PANI, J., JEFFRES, J., SHIPPEY, G., AND SCHWARTZ, K. 1996. Imagining projective transformations: Aligned orientations in spatial organization. *Cognitive Psychology. 31*, 125-167.

REINHART, W.F. 1990. *Effects of depth cues in depth judgments using a field-sequential stereoscopic CRT display*. Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA.

REISING, J.M. AND MAZUR, K.M. 1990. 3-D displays for cockpit: Where they payoff. *SPIE Proceedings. 1256*, 35-43.

SCHNEIDER, W. AND SHIFFRIN, R.M. 1977. Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review. 84*, 1-66.

SCHRIEVER, W. 1925. Experimentelle studien über das stereoskopische sehen. *Zeitschrift für Pscyhologie 96*, 113-170.

SHIFFRIN, R.M. AND SCHNEIDER, W. 1977. Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review. 84*, 127-190.

SMETS, G.J.F. 1992. Designing for telepresence: The interdependence of movement and visual perception implemented. In *Proceedings of 5th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Man-Machine Systems*, Pergamon Press.

SOLLENBERGER, R.L. 1993. *Combining depth information: Theory and implications for design of 3D displays*. Ph.D. thesis, Dept. of Psychology, Univ. of Toronto, Ontario, Canada.

SOLLENBERGER, R.L. AND MILGRAM, P. 1993. Effects of stereoscopic and rotational displays in a three-dimensional path-tracing task. *Human Factors. 35*, 483-499.

TODD, J.T. 1985. Perception of structure from motion: Is projective correspondence of moving elements a necessary condition? *Journal of Experimental Psychology: Human Perception & Performance. 11*, 689-710.

WALLACH, H. AND O'CONNELL, D.H. 1953. The kinetic depth effect. *Journal of Experimental Psychology. 45*, 205-217.

WANGER, L. 1992. The effect of shadow quality on the perception of spatial relationships in computer generated imagery. In *Proceedings of the 1992 ACM SIGGRAPH Symposium on Interactive 3D Graphics. 25*, 2 (March), ACM, New York, 39-42.

WANGER, L. FERWANDA, J., AND GREENBERG, D. 1992. Perceiving spatial relationships in computer-generated images. *IEEE Computer Graphics & Applications*. (May), 44-58.

WARE, C. AND ARTHUR, K. 1993. Fish tank virtual reality. In *Proceedings of INTERCHI'93: ACM Conference on Human Factors in Computing Systems.* ACM, New York, 37-42.

WARE, C. AND FRANCK, G. 1996. Evaluating stereo and motion cues for visualizing information nets in three dimensions. *ACM Transactions on Graphics. 15*, 121-140.

WICKENS, C.D. 1992a. *The proximity compatibility principle: Its psychological foundation and its relevance to display design*. Tech. Report ARL-92-5/NASA-92-3. Savoy, IL: University of Illinois Institute of Aviation, Aviation Research Lab.

WICKENS, C.D. AND CARSWELL, C.M. 1995. The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors. 37*, 473-494.

WICKENS, C.D. 1992b. Virtual reality and education. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, 842-847.

WICKENS, C.D., MERWIN, D.H. AND LIN, E.L. 1994. Implications of graphics enhancements for the visualization of scientific data: dimensional integrality, stereopsis, motion, and mesh. *Human Factors. 36*, 44-61.

WICKENS, C.D., TODD, S., AND SEIDLER, K. 1989. Three-dimensional displays: Perception, implementations and applications. CSERIAC Tech. Rep 89-001, Wright Patterson Air Force Base, Ohio.

YEH, Y. 1993. Visual and perceptual issues is stereoscopic display. In *Stereo Computer Graphics*, D.F. McAllister, Eds. Princeton University press, Princeton, N.J. 50-70.

YEH, Y. AND SILVERSTEIN, L.D. 1990. Visual performance with monoscopic and stereoscopic presentations of identical three-dimensional visual tasks. *SID International Symposium Digest of Technical Papers. 21*, 359-362.

YONAS, A. 1979. Attached and cast shadows. In *Perception and Pictorial Representation*, C.F. Nodine, Ed. Praeger, New York, 100-109.

YONAS, A., GOLDSMITH, L.T., AND HALLSTROM, J.L. 1978. Development of sensitivity to information provided by cast shadows in pictures. *Perception. 7*, 333-341.

YOUNG, M.J. LANDY, M.S., AND MALONEY, L.T. 1993. A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Research. 33*, 18, 2685-2696.

ZHAI, S., BUXTON, W., AND MILGRAM, P. 1996. The partial-occlusion effect: Utilizing semitransparency in 3D human-computer interaction. *ACM Transactions on Computer-Human Interaction. 3*, 3 (Sept), 254-284.