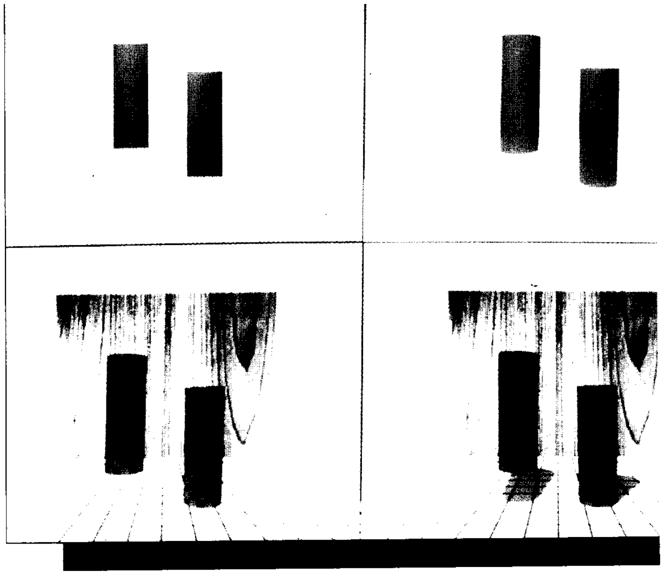


Perceiving Spatial Relationships in Computer-Generated Images

Leonard R. Wanger, James A. Ferwerda, and Donald P. Greenberg
Cornell University

Visual information determines our perception of spatial relationships. We explored the visual cues we believe are most important for depth relations in computer-generated images.

Under normal conditions our perceptions closely correspond to physical reality. Illusions result when the information our visual system relies on is absent or ambiguous.

Figure 1 illustrates the familiar Necker cube illusion. The cube spontaneously reverses its orientation with sustained viewing. In the Necker cube we see a trend evident in many illusions: 2D figures with underspecified 3D interpretations. The orthographic projection used to depict the Necker cube leaves ambiguous the relative depth ordering of the cube's front and rear faces. Faced with this ambiguity, our visual system relies on its built-in assumptions about the structure of the environment and provides us with a concrete (albeit multistable) visual experience. Observe in the Necker cube that we never see an amorphous or flat figure; the organization is always 3D, but our perceptual experience shifts between the two equally valid 3D interpretations.

The Necker cube is one of many examples that suggest we have a predisposition to interpret figures in a 3D context. For instance, we see the line junctions of Figure 2 as the corners of 3D objects. We see the smoothly varying lightness gradients of Figure 3 as bumps and dimples on a surface, and we see the varying texture gradients of Figure 4 as surfaces oriented in depth. The two spheres shown in Figure 5 have the same projected size but appear to be at different distances because of their shadows. The surfaces of the imaginary cylinders shown in Figure 6 look amorphous in this static image, but jump into relief if we rotate the cylinders. All these examples reveal the sources of information that underlie our visual experience of the 3D world around us.

0272-17-16/92/0500-0044$03.00 ©1992 IEEE

Linear perspective, surface shading, shadow and texture gradients, and the transformations of these properties with motion are sources of information that the visual system uses to construct a reliable interpretation of the shapes, sizes, positions, and orientations of objects in the environment. For reasons of difficulty or computational efficiency, computer graphics rendering schemes often approximate or omit one or more of these sources of information from the images they create. This raises the potential for illusions. These illusions might be especially insidious because the interpretations that the visual system imposes might be perfectly valid, but they might not correctly represent spatial relationships. For this reason, graphics researchers must address the following questions:

- Which sources of visual information must we have to correctly interpret spatial relations in images?
- What relative importance do different sources of visual information have with regard to metric judgments of spatial relations in images?
- How does the task in which we use the images affect the visual information's usefulness?

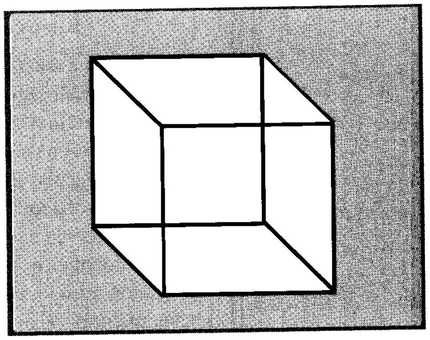


Figure 1. The Necker cube.

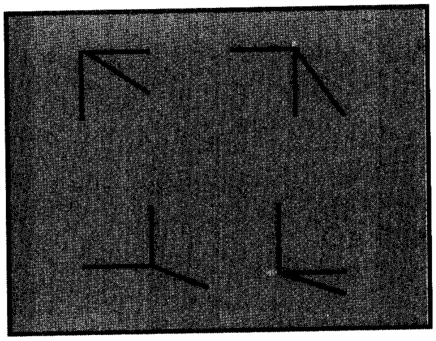


Figure 2. Corner junctions.

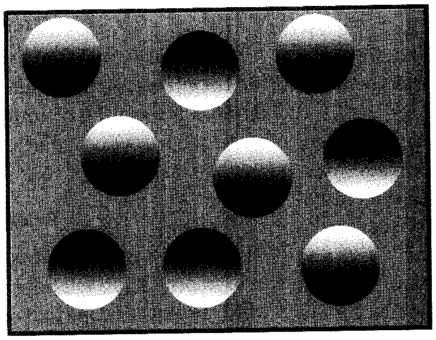


Figure 3. Surface shading (after Ramachandran's work[1]).

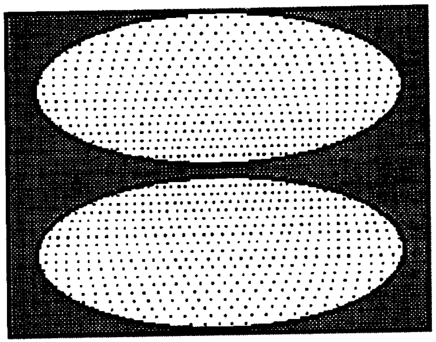


Figure 4. Texture gradients (after Gibson's work[2]).

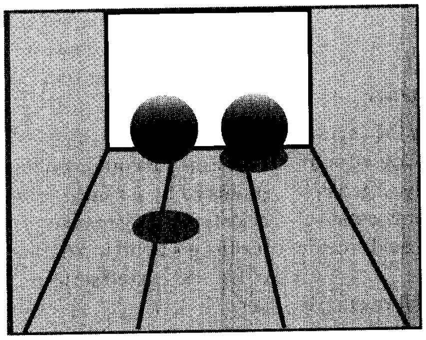


Figure 5. Shadow (after Yonas's work[3]).

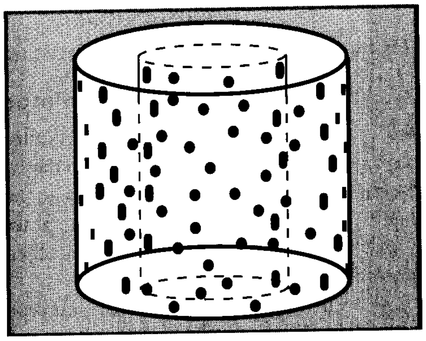


Figure 6. Relative motion (after Ullman's work[4]).

Cue theory

The study of visual perception has a rich history dating back to the ancient Greeks, who believed (much like today's ray tracers) that rays emitted from the eyes permit us to visually "touch" and experience distant objects. Much later, the light-sensitive properties of the eyes became known and interest shifted to how 2D images projected onto the retinas could inform us about the world's 3D layout.

The first problem for this approach arises because of the fundamental ambiguity in a 2D image formed by projection. A large distant object and a close one can have identical projections, so how can we perceptually determine an object's size and distance based on our retinal images? Scientists in the 1700s responded to this problem by developing the cue theory of perception,[5] which has persisted in various guises to the present day.

Cue theory states that the visual system computes the distances of objects in the environment based on information from the posture of the eyes and from the patterns of light projected onto the retinas by the environment. There is ongoing controversy over the specific forms of the cues, whether they are innately known or learned, and whether they directly specify

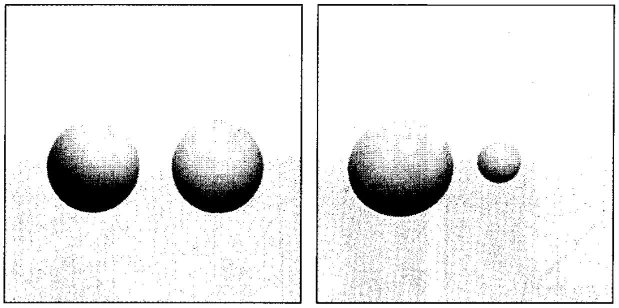


Figure 7. The effects of orthographic and perspective projections.

spatial properties or are used to infer them. In the following discussion, we outline the commonly accepted aspects of this approach.

Psychologists distinguish between two fundamental classes of cues. The so-called primary cues achieve their primacy by ostensibly providing a physiological basis for the direct measurement of distance. The second class of cues, pictorial cues, are so called because they were first observed by graphic artists and painters who used them to create the impression of depth in pictures. However, we now know that many of these cues are not just trompe l'oeil tricks but fundamental sources of information for visual perception.

Primary cues

First we will describe two of the primary cues.

Convergence and accommodation

Early cue theorists proposed that by registering the angle of convergence between the two eyes when they are focused on an object, viewers could determine distance to the object geometrically. Others thought the changing state of accommodation of the eye's lens to focus on near and distant objects could provide information about an object's distance.

Possibly the visual system uses these sources of information in perception, but it is unlikely that they are physiological, because researchers have not shown the necessary feedback systems from the muscles of the eyes to the visual pathways.[6]

Binocular disparity

A powerful cue that falls in the primary class is binocular disparity. It is well known that projective differences between the two retinal images provides the basis for stereo vision, which allows us to make our finest depth judgments. But binocular disparity alone cannot account for our experience of depth because we can correctly perceive depth relations when no disparity is present. This occurs when viewing objects at distances greater than a few meters,[7] under monocular viewing conditions, and when viewing pictures, where depth relations are represented on the picture's flat surface.

Pictorial cues

While we included a discussion of the primary cues here for completeness, our interest in these experiments is really directed toward the pictorial cues because they provide the basis for our understanding of spatial relationships in computer graphics imagery.

Perspective

Perspective projection has been used since the Renaissance to represent depth relations.[8] It is a geometric construction in which an object's depicted size varies inversely with its distance from the center of projection. Perspective establishes a size/distance relation in which, if we know the actual size of an object, we can compute its distance and vice versa. If neither size nor distance is known a priori, we cannot resolve the projective ambiguity by perspective information alone. Figure 7 compares the orthographic and perspective projections of a scene.

Texture

Gibson[2] first proposed that a viewer uses the visual gradient produced by a slanted textured surface as information for its orientation. Experiments have since confirmed that we can use texture gradients to determine the spatial orientation and relief of flat and curved surfaces.[9] Figure 8 illustrates these properties.

Gibson also suggested that texture reveals spatial relations by enhancing the information that one surface occludes another, particularly in the case of moving objects, where texture elements are successively hidden and revealed at occluding boundaries.[10]

Shading and shadow

Similarly to gradients of texture, gradients of illumination can visually specify the shape and orientation of a surface.[11] Shadows, caused by one object

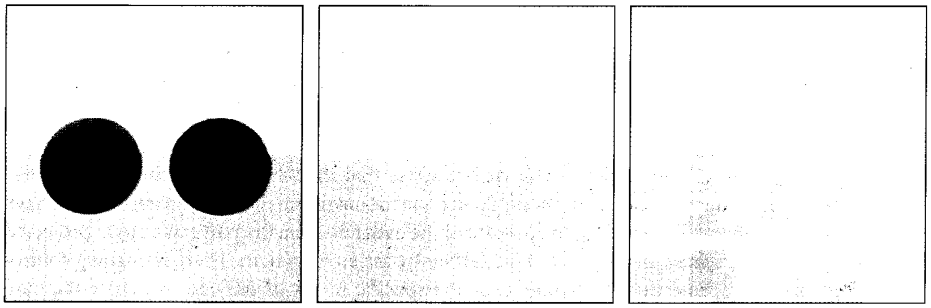


Figure 8. The effects of texture gradients on perceived shape and orientation.

occluding the light falling on another, are a powerful source of information for spatial position.[3] Because gradients of illumination can specify both shape and location, it helps to distinguish between the effects of shading and shadow shown in Figure 9.

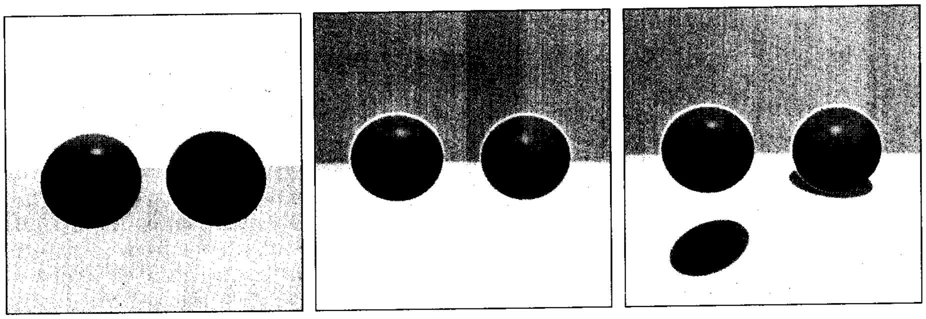


Figure 9. The effects of shading and shadow on perceived shape and spatial position.

Motion

The relative displacement between objects at different distances caused by the motions of objects or by movement of the point of observation has long been noted as a source of information for spatial relations.[12] Ullman[4] showed that the relative motions of points on an object can specify object geometry. But as others[13,14] have pointed out, relative motion information in isolation does not resolve depth order and directional ambiguities. However, motion in the presence of other cues like perspective projection or textural occlusions can specify spatial relations powerfully. Figure 10, for example, shows three frames from a sequence simulating movement of the point of observation in a scene.

Reference frames

We typically see objects as elements in a bounded environment. In the absence of an environmental frame of reference, we often confuse an object's size and distance. Witness the moon illusion, in which the perceived size of the lunar disk changes drastically as it moves from horizon to zenith. At the horizon, we can reference the moon in size and distance to other objects on the ground plane. At the zenith, there are no such reference points, and the moon's size and distance are confounded.[15] Environmental reference frames also play themselves out in the so-called equidistance tendency,[16] in which an object tends to be drawn, in perceived distance, to the distance of its surround.

Cues selection

Over the years perception scientists have suggested many other cues for spatial relations. In our studies we were sometimes unable to examine particular cues because of technical limitations (lack of stereo display hardware, lack of eye tracking equipment). In other cases we omitted cues because they seemed too specialized (aerial haze) or cognitively mediated (familiar size). In the end we arrived at a set of test cues (listed below) that we believe represents the most important sources of visual information for depth relations in computer-generated images. This set also contains the classes of visual information computed by a broad range of rendering algorithms.

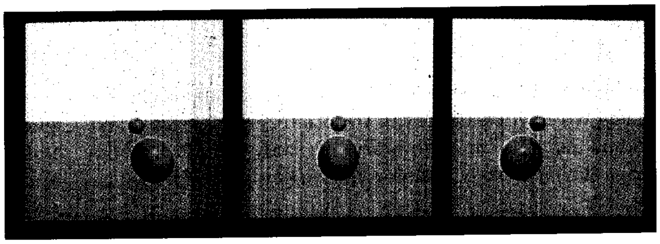


Figure 10. The effect of relative motion as the point of observation moves from left to right.

Experiments

In three experiments we assessed the influence of pictorial cues on perceived spatial relations in computer-generated images. Each experiment examined the accuracy with which subjects matched the position, orientation, and size of a test object with a standard by interactively translating, rotating, and scaling the test object. The following sections describe the pictorial cues tested and the general format of the experiments. The specific details of the objects, environment, methods, and procedure for all the experiments are described in the sidebar "Experimental setup."

Cues tested

We examined six pictorial cues in the experiments:

- Projection: Displays used either orthographic or perspective projection.
- Shadow: We rendered shadows by projecting shadow-casting objects onto the ground plane.[17] Shadows were either present or absent in each experimental trial.
- Object texture: In all displays showing the object texture cue, we texture mapped a uniform checkerboard pattern onto each object in the scene. In each experimental trial, objects were either textured or solid colored.
- Ground texture: In all displays showing the ground texture cue, we applied a randomly perturbed checker pattern to the ground plane and the back wall of the environment. This provided statistically uniform texture gradients without allowing subjects to count the checks to measure distances. In each ex-

Experimental setup

Subjects: The same 12 subjects participated in each experiment. All subjects were graduate students or staff members at Cornell's Program of Computer Graphics and had significant experience viewing computer-generated imagery of 3D environments. All subjects had normal or corrected to normal vision.

Apparatus: Images were generated in real time on a Hewlett-Packard 9000 series 835 workstation with a Turbo SRX accelerated 3D graphics subsystem. We displayed them on a Hewlett-Packard 98752A 19-inch color monitor. The active image region was approximately 13 inches wide and 10.5 inches high, with a resolution of 1,280 × 1,024 pixels. Subjects viewed the monitor under normal room lighting from a distance of approximately 18 inches. The display subtended 19.8 degrees of visual angle horizontally and 16.26 degrees of visual angle vertically. The experimenter set contrast and brightness and held them constant for all subjects. Viewing was binocular, and head and body movements were not restricted.

Displays: We rendered all images using a Phong illumination model[1] and Gouraud shading[2] on a 12-bit color, dithered, double-buffered display. The images depicted objects in a virtual room whose geometry coincided with the geometry of the monitor. Thus the back wall of the virtual room coincided with the back of the monitor, and the ground plane of the virtual room coincided with the bottom of the monitor. We chose to have the virtual space coincide with the monitor space to ensure that the perspective projection used to depict the virtual room was correct for the 18-inch viewing distance used in the experiments.

An on-going debate in the psychology literature argues the perceptual consequences of viewing images from other than their station points. We chose the conservative opinion and made viewing distance and projective distance coincident in our displays.[3]

We set the virtual camera position used in rendering at the eyepoint of the observer looking toward the center of the room, with the frustum of view coinciding with the physical space of the monitor. This yielded a view of the room such that only the back wall and ground plane of the room were visible at any time. The edges of the back wall and ground plane lay outside the field of view.

Objects were illuminated by an ambient light (RGB emission values of 0.6, 0.6, 0.6) and by a point light source (RGB emission values of 0.6, 0.6, 0.6) from a position in the virtual space above the subject's right shoulder (6.5 inches to the right of center on the screen, 39 inches above the screen, and 26 inches away from the screen). This position provided natural-looking overhead illumination with good specular highlights and good shaded definition of the objects.

In each experiment the objects in the virtual room lay on an imaginary line segment rotated by random values around the *y* (vertical) and *z* (depth) axes on a trial by trial basis. The range of rotation was 21.5 through 50.0 degrees for both the *y* and *z* axes, randomly set to either clockwise or counter-clockwise. We added a second rotation to keep the orientation of the objects constant from trial to trial. The imaginary line subtended 5 degrees of visual angle on the screen and allowed simultaneous viewing of the objects without eye movements.

Procedure: In all the experiments, subjects pressed a button to indicate satisfaction with their solutions to the task. Pressing the button recorded the state of the test object, blanked the screen, and called up the next trial.

Subjects first worked through a block of practice trials, which included all the cues in each of their states at least once. Following the practice block, we presented subjects with randomly ordered test trials consisting of a single repetition of all combinations of the cues. The test trials were presented in a single session for the positioning task, and over two sessions for the rotation and scaling tasks. Some subjects went through all the trials in a single session for the scaling task. Each session lasted approximately 30 minutes. The exact amount of time per session for all of the experiments depended on the speed of the individual subject.

References

1. R. Hall, *Illumination and Color in Computer-Generated Imagery*, Springer-Verlag, New York, 1989.
2. H. Gouraud, "Continous Shading of Curved Surfaces," *IEEE Trans. Computers*, Vol. C-20, No. 6, June 1971, pp. 623-629.
3. J. Farber and R.R. Rosinski, "Geometric Transformation of Pictured Space," *Perception*, Vol. 7, No. 2, Feb. 1978, pp. 269-282.

perimental trial, the walls were either texture mapped or solid gray. (We divided the texture cue into object texture and ground texture in order to separately study the effects of the shape-defining and occlusion-enhancing properties of the texture cue.)

- Motion: We provided motion by moving the eye point along a horizontal axis while rotating the view vector to keep the fixation point constant. This allowed subjects to see a 20-degree range of views around the vertical axis. Sinusoidal accelerations provided natural looking motion. In each experimental trial, the viewpoint was either moving or stationary.
- Elevation: In the scaling and rotations tasks (Experiments 2 and 3 below), we tested the effect of a frame of reference on size/distance relations. In each experimental trial, the test object either rested on the ground plane or floated in the air.

Experiment 1: Positioning task

Experiment 1 examined the ability to determine an object's 3D position using images containing different combinations of pictorial cues.

The display depicted three balls in a room. Two of the balls were fixed in position, with one ball lying on the ground plane and the other floating in the air. The two fixed balls formed the endpoints of an imaginary line segment whose orientation varied randomly from trial to trial. The third ball, initially located at a random position in space, could be moved by adjusting three knobs. Each knob moved the ball along one of the principal axes of the environment. We restricted the extent of the motion to prevent the balls from interpenetrating or occluding one another.

In each trial of the experiment, subjects were instructed to move the adjustable ball to lie at the midpoint of the imaginary line segment joining the two outer balls. Figure 11 shows the positioning task with all of the pictorial cues present.

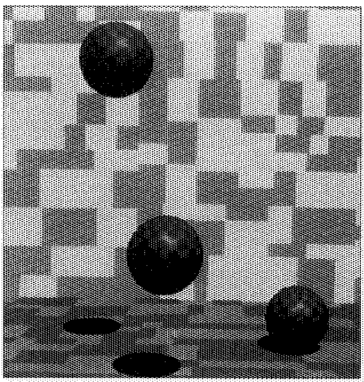


Figure 11. The positioning task.

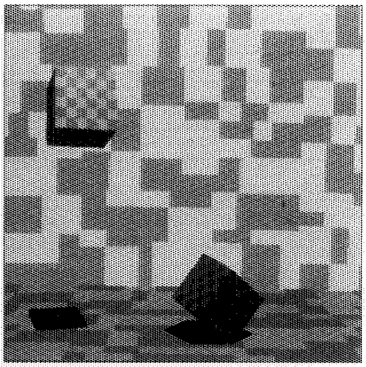


Figure 12. The rotation task.

Experiment 2: Rotation task

Experiment 2 examined the ability to determine an object's 3D orientation using images containing different combinations of pictorial cues.

The display depicted two cubes in a room. One cube lay on the ground plane and the other cube floated in the air. In addition, one of the cubes was fixed in orientation; the other could be rotated by adjusting three knobs. Whether the fixed cube floated in the air or lay on the ground plane depended on the state of the elevation cue. The orientation of the fixed cube varied randomly from trial to trial. Each knob rotated the test cube around one of its principal axes.

In each trial, subjects were instructed to adjust the movable cube until its orientation matched the orientation of the fixed cube. Figure 12 shows the rotation task with all the pictorial cues present.

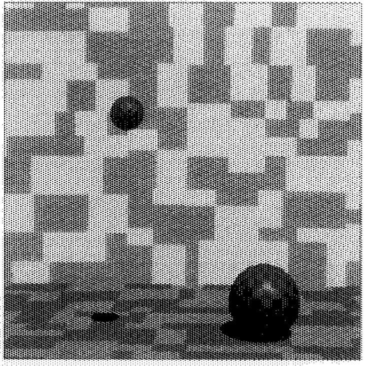
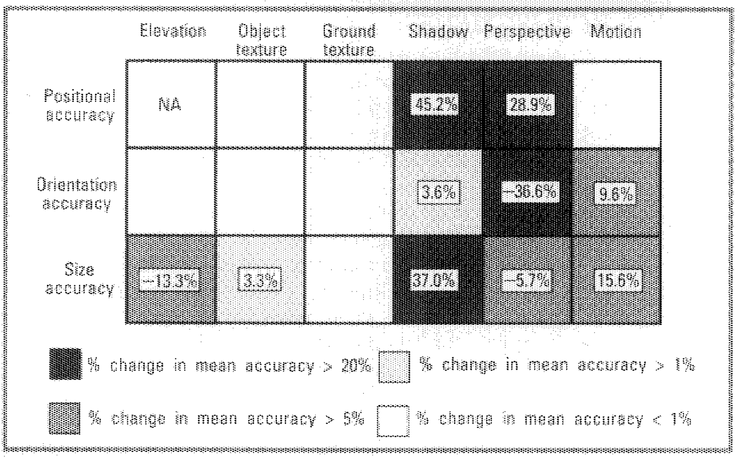


Figure 13. The scaling task.



	Elevation	Object texture	Ground texture	Shadow	Perspective	Motion
Positional accuracy	□ NA	□	□	■ 45.2%	■ 28.9%	□
Orientation accuracy	□	□	□	░ 3.6%	■ −36.6%	▨ 9.6%
Size accuracy	▨ −13.3%	░ 3.3%	□	■ 37.0%	▨ −5.7%	▨ 15.6%

■ % change in mean accuracy > 20%
░ % change in mean accuracy > 1%
▨ % change in mean accuracy > 5%
□ % change in mean accuracy < 1%

Figure 14. Matrix of significant results for the experiments.

Experiment 3: Scaling task

Experiment 3 examined the ability to determine an object's 3D size using images containing different combinations of pictorial cues.

The display depicted two balls in a room. One ball lay on the ground plane, and the other ball floated in the air. In addition, one of the balls was fixed in size, while the other could be scaled by adjusting a knob. Whether the fixed ball floated in the air or lay on the ground plane depended on the state of the elevation cue. The initial sizes of the balls and their distances from the viewpoint varied randomly from trial to trial. The knob scaled the adjustable ball equally on each of its principal axes.

In each trial, subjects were instructed to scale the adjustable ball until its size matched the size of the fixed ball. Figure 13 shows the scaling task with all pictorial cues present.

Results

Figure 14 shows the results of the first three experiments. The cues that significantly affected performance accuracy in a task are shaded according to the magnitude of their effect. The number shown in each box represents the percentage change in mean accuracy that the cue produced in a specific task. A positive percentage indicates that the cue increased accuracy, and a negative percentage indicates that the cue decreased accuracy.

The following sections elaborate on the experimental results. A detailed description of the methods used to determine the relative importance and direction of effect of the cues appears in the sidebar "Statistical methods." Detailed results of all the experiments are presented in the sidebar "Experimental results" at the end of the article.

Cues and positional accuracy

Cues had the following effects on positional accuracy:

- Shadow had a dominant effect, increasing accuracy by 45.2 percent and reducing the mean positioning error from 0.546 inches to 0.299 inches. As a point of reference, the ball's radius was 0.55 inches (see Figure 15).
- Perspective projection increased accuracy by 28.9 percent, reducing the mean error from 0.494 inches to 0.351 inches (see Figure 16).
- Motion, object texture, and ground texture did not significantly affect positional accuracy.

The positioning task requires knowledge of the relative locations of the balls. This is inherently a spatial task because subjects cannot perform it by simply comparing the projected sizes or shapes of the balls. Instead, they must infer or assume the spatial positions of the balls. Shadow allows subjects to infer position because the separation between a ball and its shadow indicates the ball's height above the ground plane, and the shadow's location on the ground plane indicates the ball's distance.

Perspective allows subjects to infer position by providing a size/distance gradient by which they can determine the balls' relative positions if they know their sizes or assume them to be constant and equal.

Statistical methods

The analysis of variance is a common method of separating the effects of multiple factors in a data set. The ANOVA examines which factors have a significant influence on a dependent variable by comparing the variance within a factor to the variance between factors. In the experiments described in this article, the factors are the pictorial cues, and the dependent variable is a measure of accuracy for the task.

The multivariate analysis of variance, or MANOVA, is an extension to the ANOVA method that accounts for multiple dependent variables. The repeated measures MANOVA is a further extension of the ANOVA method that accounts for multiple measures on the same experimental unit (in this case, multiple trials with the same subject).

The ANOVA calculates two important statistics for each factor: the *F* value and its associated *p* value. In a set of observations, the *F* value is a measure of the variation caused by a particular experimental factor after all other factors have been taken into account. We can think of it as an estimate of how well a factor accounts for the behavior of a dependent variable. The *p* value gives the significance probability associated with the *F* value. It is the probablility that the amount of variation seen for the factor in the data could have arisen merely by random variation. A *p* value of 0.1 means that there is a 1 in 10 chance that the statistical result seen is due to random variation in the data. We require a *p* value of $p < 0.05$ (a 1 in 20 chance) for a factor to be considered a significant factor. A significance level of $p < 0.05$ is an accepted significance criterion in perception psychology.

Single factors that have a significant effect on the dependent variables are called main effects. In addition to the main effects, the ANOVA also tests for higher order interactions between factors. Higher order interactions, also known as crossed effects, are factors whose effects depend on the state of one or more additional factors.

Once the multivariate ANOVA reveals which factors have a significant effect on a dependent variable, we can use inferential techniques to determine the direction of the effect. In the case of the experiments we described here, the direction of the effect shows whether the accuracy of performance on the task increased or decreased.

The inferential method we used is the matched pairs *t* test. A matched pairs *t* test compares the mean values for experimental observations with and without a particular factor. The sign of the *t* test tells whether the factor had a positive or negative effect. The *t* test also produces an associated *p* value that indicates how reliably different the tested means are. This *p* value and the *p* value produced from the *F* statistic will differ because of the number of independent observations, or degrees of freedom (*df*), that go into calculating the *t* and *F* statistics. In general, the *F* statistic is a more conservative and powerful measure of an effect, so its *p* value will be used as an indication of an effect's reliability.

Finally, it is useful to determine which cues are most important in each task. We can weight the cues by determining the percentage of the total variation in accuracy that a factor accounts for in a task. We can then use the relative weightings to rank the factors in order of their importance.

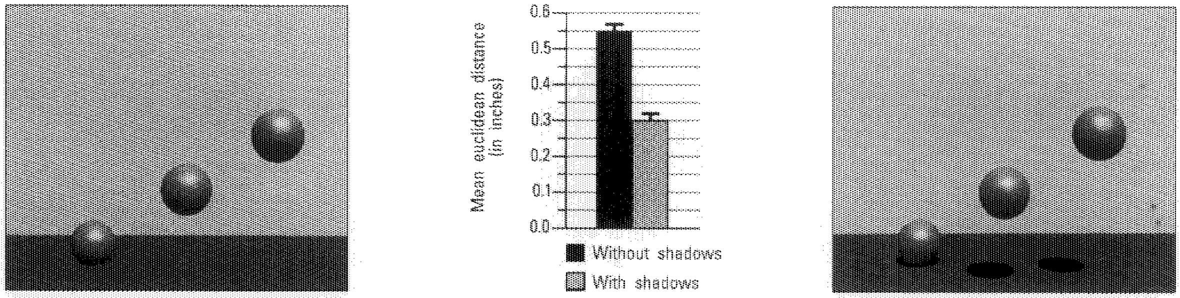


Figure 15. Results for the shadow cue in the positioning task.

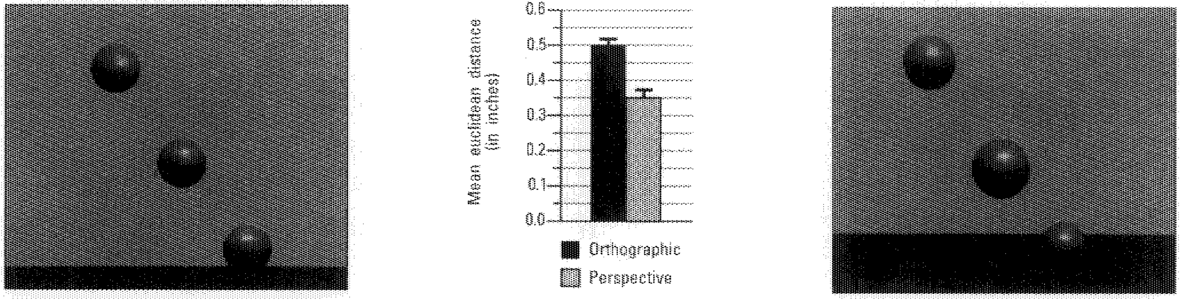


Figure 16. Results for the perspective cue in the positioning task.

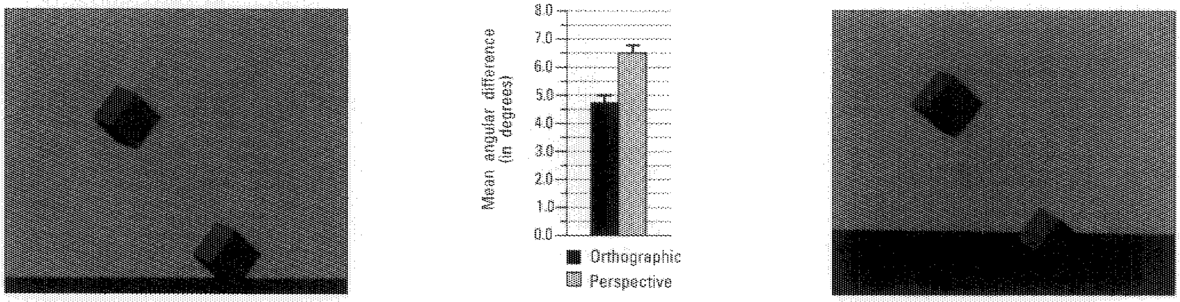


Figure 17. Results for the perspective projection cue in the orientation task.

Since we know motion gradients are an important source of information about objects' relative positions, it seems curious that motion did not significantly affect accuracy in the positioning task. We believe that motion would have been a significant cue if our implementation had been less confusing. Subjects reported that they were disturbed by the forced motion of the viewpoint used in the experiment. This probably counteracted any advantage the motion cue might have provided.

Cues and orientational accuracy

Cues had the following effects on orientational accuracy:

- Perspective projection had a dominant effect, decreasing performance accuracy by 36.6 percent. This increased the mean orienting error from 4.811 degrees to 6.574 degrees (see Figure 17).
- Motion increased performance accuracy by 9.6 percent, decreasing the mean error from 5.979 degrees to 5.406 degrees (see Figure 18).
- Shadows had a small effect, increasing accuracy by 3.6 percent. Mean error was decreased from 5.797 degrees to 5.588 degrees (see Figure 19).
- The object texture, ground texture, and elevation cues did not significantly affect accuracy in the orienting task.

If we use orthographic projection, subjects can accomplish the orienting task by simply matching the projected shapes of the cubes. When we use perspective projection, this simple relation no longer holds, and subjects must determine the cubes' actual 3D orientations. This accounts for the detrimental effect of perspective projection in the orienting task. Note that even when the cubes actually align, they still appear out of alignment in the perspective image because they occupy different spatial locations (see Figure 17).

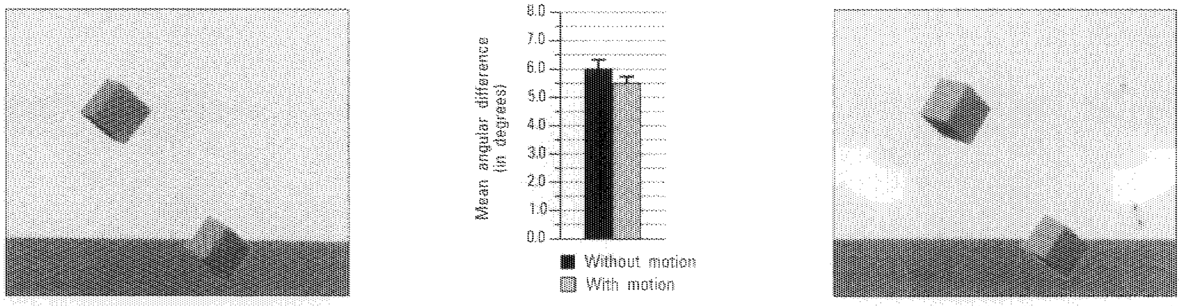


Figure 18. Results for the motion cue in the orientation task.

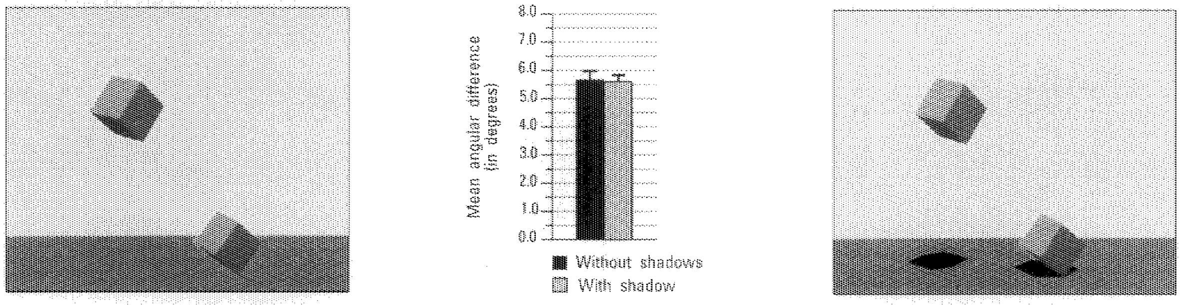


Figure 19. Results for the shadow cue in the orientation task.

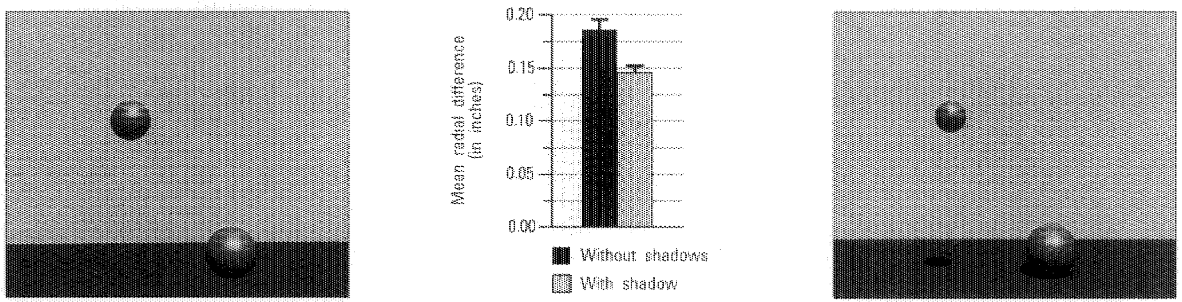


Figure 20. Results for the shadow cue in the scaling task.

Motion facilitates accurate orienting by revealing a transformational equality between the two cubes. Because the axis of rotation is centered between the cubes, the ratios of horizontal displacement of corresponding pairs of points on the cubes are equal when the cubes align. Observers perceiving this equality could use it to facilitate the orienting task.

The shadow cue had a small positive effect. This likely occurred because shadows reinforced the projected-shape-matching strategy subjects used in the orthographic trials. In trials with shadows and orthographic projection, subjects could use a match between the shadows' shapes as a secondary check for a match between the cubes' projected shapes. The magnitude of the effect was small because only one fourth of the trials had both shadows and orthographic projection.

Cues and size scaling accuracy

Cues had the following effects on size scaling accuracy:

- Shadow had a dominant effect, increasing performance accuracy by 37.0 percent. Shadow decreased the mean radial scaling error from 0.185 inches to 0.147 inches (see Figure 20). As a point of reference, the radius of the standard ball ranged from 0.514 inches to 0.615 inches.
- Motion increased performance accuracy by 15.6 percent, decreasing the mean error from 0.174 inches to 0.161 inches (see Figure 21).
- Elevation decreased accuracy of performance by 13.3 percent, increasing the mean error from 0.163 inches to 0.173 inches (see Figure 22).
- Perspective projection decreased accuracy of performance by 5.7 percent, increasing the mean error from 0.166 inches to 0.171 inches (see Figure 23).
- Object texture increased accuracy of performance by 3.3 percent, decreasing the mean error from 0.169 inches to 0.166 inches (see Figure 24).

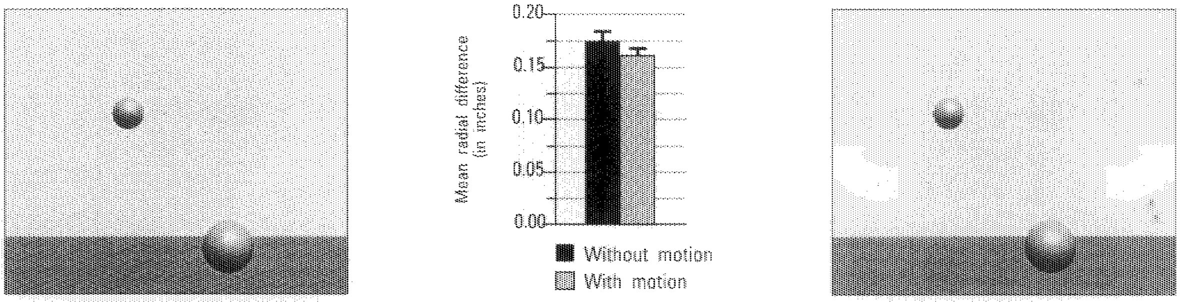


Figure 21. Results for the motion cue in the scaling task.

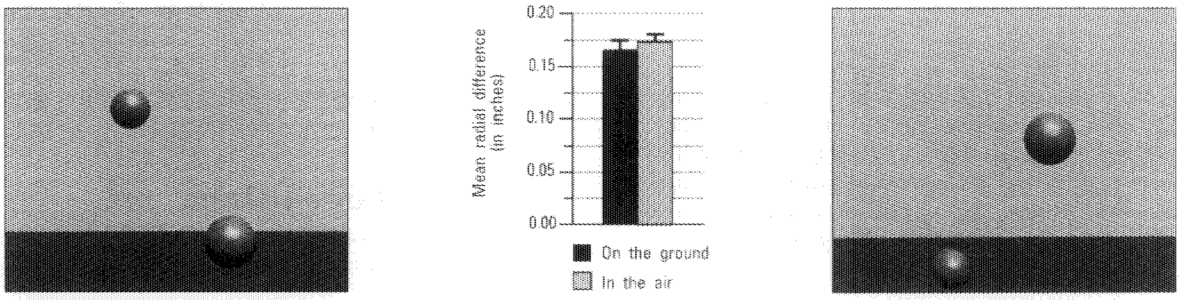


Figure 22. Results for the elevation cue in the scaling task.

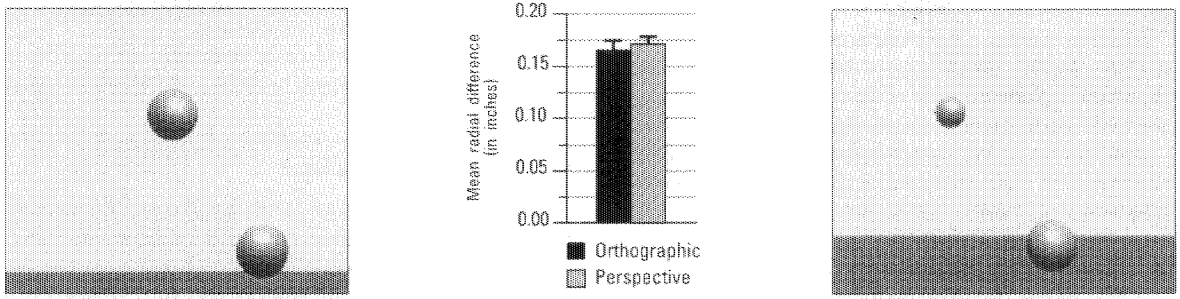


Figure 23. Results for the perspective projection cue in the scaling task.

- The ground texture cue did not significantly affect accuracy in the scaling task.

The scaling task incorporates aspects of both the orienting and positioning tasks. Like the orienting task, simple matching of projected size is effective when the trial utilizes orthographic projection. However, like the positioning task, subjects must know the balls' spatial locations to accomplish the task under perspective projection.

The shadow cue was effective for reasons similar to those given in both the positioning and orienting tasks: It defines spatial location in perspective trials and reinforces projected size matching in orthographic trials.

Because projected size is invariant with motion under orthographic projection, it is likely that the increase in accuracy seen with motion results solely from the location-defining property of motion gradients in the perspective trials. Probably the motion cue helped here because the property of the object being adjusted (size) was not confounded by the motion, as occurred in the positioning task.

The presence of object texture increased accuracy in the scaling task by a small amount. We do not have a conclusive explanation for this effect, but we believe object texture enhances the perceived relief of the balls. This could facilitate the task by more firmly anchoring the balls in a 3D context.

Because the scaling task directly addresses size/distance relations, it is logical that the frame-of-reference cue, as represented by elevation, plays a role in the task. We can attribute the decrease in accuracy seen when the ball was elevated to the loss of the reference frame provided by the ground plane.

Because the scaling task concerns size/distance relations, it makes sense that both the cues that specify position and the cues that specify size are active. The analysis showed that the bulk of the variation in accuracy resulted from interactions

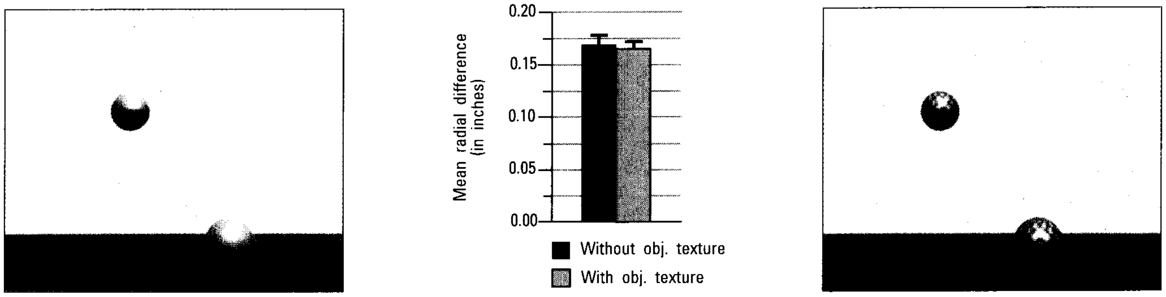


Figure 24. Results for the object texture cue in the scaling task.

between the cues. Although the effects of the individual cues are valid, they represent only part of the complex pattern of results seen in the scaling task. Probably this pattern of interactions between cues will be common to many spatial manipulation tasks because they typically involve changing both the intrinsic properties of objects and their locations.

Caveats

While our experiments provided results suggesting that visual cues are differentially effective in different tasks, we think it is important to point out some of the complexities of our experimental design that temper the conclusions we will draw.

Simultaneous testing of many cues allowed us to compare the magnitudes of their effects, but it also caused an exponential explosion in the number of trials necessary for each subject. To keep the trials per subject down to a manageable number, we randomly varied the initial positions of the objects in the environment instead of testing each experimental condition at a fixed set of positions. Using a fixed set of positions would have been advantageous for two reasons. First, it would have increased the number of measures taken for each condition, which would increase the statistical reliability of the results. Second, it would have simplified the statistical analysis of the data.

A factor we held constant but probably should have randomized was the presentation order of the tasks. The fact that all the subjects performed the experimental tasks in the same order—positioning, orienting, then scaling—might have influenced the strategies they used.

Additionally, two aspects of the experiments' implementation might have adversely affected the subjects' performance. First, subjects did not like using knobs as input devices for the tasks. These 3D tasks really need input devices that allow independent simultaneous adjustment of multiple spatial parameters. Second, we feel certain that relative motion is more important for specifying spatial relations than these experiments indicate. We believe that much of motion's effectiveness was counteracted by our forcing it on the subjects in our experiments. Many subjects reported that forced motion of the viewpoint was very disturbing. The motion cue would probably be more effective if we put it under user control.

Conclusions

At the outset of these investigations we posed three questions concerning the efficacy of pictorial information in conveying spatial relations:

- Which sources of visual information must we have to correctly interpret spatial relations in images?
- What relative importance do different sources of visual information have with regard to metric judgements of spatial relations in images?
- How does the task in which we use the images affect the visual information's usefulness?

We hoped that one or two of the pictorial cues would dominate in all the tasks. Then we could tell the graphics community that including these cues in their displays would much improve performance in spatial manipulation tasks. But like most issues concerning human perception and performance, the situation is more complicated than it first appears. While a few cues did account for most of the changes in performance seen in the tasks, the effectiveness of different cues and the directions of their effects varied with the task.

In the positioning task, where subjects needed object location information to perform the task, shadow and perspective were significant cues. Shadow provided a ground-plane-relative reference for height and distance, and perspective provided a size/distance gradient.

In the orienting task, where object location was unimportant but subjects needed information about the relative orientations of the object's faces, perspective, motion, and shadow were the significant cues. However, in this case, the presence of perspective projection hurt performance by preventing subjects from performing the task by simply matching the projected shapes of the objects. Motion was an effective cue, revealing a transformational equality when the objects were in the same orientation. Shadow was a minor cue, providing redundant information for the shape-matching strategy used under orthographic projection.

Finally, in the scaling task, where both object location information and information about the object's intrinsic size were relevant to the task, all the above-mentioned cues were effective singly. But more importantly, we found that interactions between cues accounted for most of the variation in performance seen in the task.

These results suggest that, in complicated geometric modeling and spatial manipulation tasks where the 3D positions, orientations, sizes, and shapes of objects vary, we must pay careful attention to the information required to accomplish the task so that we can provide appropriate visual cues. These findings point toward an adaptive approach to user interface design for interactive 3D computer graphics in which the visual information provided in the displays changes in concert with the task at hand.

Future work

While our experiments suggest a new framework for developing interactive 3D displays, the framework is far from complete. Our experiments show that our ability to perform spatial manipulation tasks based on computer-generated images is a function of both the nature of the task and the visual information presented in the images. This observation provides two clear directions for future research. First, we need to develop a taxonomy of tasks that classifies the types of spatial information needed to perform a task and the visual cues that provide that information. Second, we need to test a broader range of visual cues for spatial relations in order to determine the kinds of spatial information they provide. Clearly, a useful test would be to compare the efficacy of the pictorial cues against a baseline given by binocular stereo vision.

To move from the limited experimental results presented here to a set of guidelines for designing 3D interactive displays, we must investigate two additional areas. First, we need to study typical applications involving 3D interaction, like geometric modeling and computer-aided design, to understand how users interact with the space represented in images. Second, we need to understand how patterns of interaction change as users become more sophisticated and as their goals and strategies change in response to the demands of specific tasks ❑

Acknowledgments

Special thanks go to the staff and students at Cornell's Program of Computer Graphics for volunteering as subjects. Additionally, we would like to thank Kathleen Dorries and Richard Darlington of Cornell's Department of Psychology for their help with the statistical analysis of the experimental data and James Cutting of Cornell's Department of Psychology for help with the experimental design and analysis.

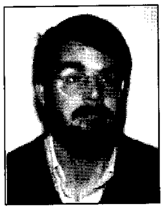
The experiments were run on equipment donated by Hewlett-Packard and Digital Equipment. This work was conducted under National Science Foundation Grant No. DCR8203979—"Interactive Input and Display techniques."

References

1. V.S. Ramachandran, "Perceiving Shape from Shading," *Scientific American*, Vol. 258, No. 2, Aug. 1988, pp. 58-65.
2. J.J. Gibson, *The Perception of the Visual World*, Houghton Mifflin, Boston, Mass., 1966.
3. A. Yonas, "Attached and Cast Shadows," in *Perception and Pictorial Representation*, C.F. Nodine, ed., Praeger, New York, 1979, pp. 100-109.
4. S. Ullman, *The Interpretation of Visual Motion*, MIT Press, Cambridge, Mass., 1979.
5. G.A. Berkeley, *A New Theory of Vision and Other Writings*, Dutton, New York, 1922.
6. A. Skavensi and R. Steinman, "Control of Eye Position in the Dark," *Vision Research*, Vol. 10, No. 1, Jan. 1970, pp. 193-203.
7. C. Graham, *Vision and Visual Perception*, Wiley, New York, 1965.
8. I. Carlbom and J. Paciorek, "Planar Geometric Projections and Viewing Transformations," *Computing Surveys*, Vol. 10, No. 4, Dec. 1978, pp. 465-502.
9. J.E. Cutting and R.T. Millard, "Three Gradients and the Perception of Flat and Curved Surfaces," *J. Experimental Psychology: General*, Vol. 113, No. 2, Feb. 1984, pp. 198-216.
10. J.J. Gibson, *The Senses Considered as Perceptual Systems*, Houghton Mifflin, Boston, Mass., 1966.
11. J.T. Todd and E. Mingolla, "Perception of Surface Curvature and Direction of Illumination from Patterns of Shading," *J. Experimental Psychology: Human Perception and Performance*, Vol. 9, No. 4, April 1983, pp. 583-595.
12. H. von Helmholtz, *Handbook of Physiological Optics*, Optical Society of America, New York, 1925.
13. J. Hochberg, *Perception*, Prentice Hall, Englewood Cliffs, N.J., 1978.
14. J. Farber and A. McConkie, "Linkages Between Apparent Depth and Motion in Linear Flow Fields," *Bulletin of the Psychonomic Society*, Vol. 10, No. 2, Feb. 1977, p. 250.
15. F. Restle, "Moon Illusion Explained on the Basis of Relative Size," *Science*, Vol. 167, No. 3,901, April 5, 1970, pp. 1,092-1,096.
16. W. Gogel, "The Organization of Perceived Space," *Psychologisch Forschung*, Vol. 36, No. 2, Feb. 1973, pp. 195-247.
17. J.F. Blinn, "Me and My (Fake) Shadow," *IEEE CG&A*, Vol. 8, No. 1, Jan. 1988, pp. 82-86.

Leonard R. Wanger is a senior engineer at SimGraphics Engineering. His research interests include perceptual issues in image synthesis, real-time image generation techniques, and user interfaces for interaction with virtual environments.

Wanger received his bachelor's degree in computer science in 1987 from the University of Iowa and his master's degree in computer graphics in 1991 from the Program of Computer Graphics at Cornell University. He is a member of ACM Siggraph and IEEE.



James A. Ferwerda is a member of the research staff at the Program of Computer Graphics at Cornell University. His research interests include perceptual issues in image synthesis, multimedia information systems, and computer-supported cooperative work.

Ferwerda received his bachelor's degree in experimental psychology in 1980 and his master's degree in computer graphics in 1987, both from Cornell University. He is a member of ACM Siggraph, IEEE, and SMPTE.

Donald P. Greenberg is the Jacob Gould Shurman Professor of Computer Graphics at Cornell University, where he holds a joint appointment in the College of Architecture and the College of Engineering. He received the ACM Siggraph Steven A. Coons Award in 1987 and the NCGA academic Award in 1989. In 1991 he was inducted into the National Academy of Engineering.

Greenberg serves as director of the Cornell Program of Computer Graphics and as director of the National Science Foundation Science and Technology Center for Computer Graphics and Scientific Visualization. He is a member of ACM Siggraph, IEEE, and Eurographics.

Readers may contact Wanger at SimGraphics Engineering, 1137 Huntington Dr. #7, S. Pasadena, CA 91030. Ferwerda and Greenberg can be reached at Program of Computer Graphics, 580 Theory Center, Cornell, University, Ithaca, NY 14853.

Experimental results

Experiment 1: Positioning

We calculated the Euclidean distance from the center of the test ball to the midpoint of the line joining the two outer balls as a measure of positional accuracy in each trial. We ran a MANOVA (see the sidebar "Statistical methods") on the measures, with the pictorial cues serving as repeating factors. We found two main effects: shadow and projection (see Table 1). Object texture, ground texture, and motion did not have statistically significant effects in Experiment 1 (all had $p > 0.3$). Additionally, there were several higher order interactions (see Table 2).

Because the fixed balls were rotated by randomly varying angles in the *y* and *z* axes, a confound exists in the data. In addition to the presence or absence of the five spatial cues, each trial was characterized by a unique arrangement of the balls in space. We ran a second ANOVA to separate the effects of the ball position and the spatial cues. In this analysis, the *y* and *z* axis rotations describing the ball positions and the five spatial cues tested in the above analysis were regressed on the Euclidean distance. The analysis revealed that the *y* and *z* axis rotations had no significant effect (both had $p > 0.1$). We had to use the ANOVA for this analysis, as the data (with the inclusion of the *y* and *z* rotations) did not fit the format of the repeated measures MANOVA. However, the univariate test estimates of statistical significance are valid, as our repeated factors have only two levels. The univariate and multivariate methods give identical results when the repeated measures have only two levels.[1]

We ran post hoc analyses in the form of matched pairs *t* tests to determine the direction of the main effects for the positioning task. The matched pairs *t* tests revealed that the presence of shadows and the presence of perspective projection significantly reduced the Euclidean distance ($t = 8.050$, $df = 382$, $p < 0.001$ and $t = 4.092$, $df = 382$, $p < 0.001$, respectively). Table 3 shows the mean accuracies for the main effects. As a point of reference, the width of the balls used in this experiment was 1.1 inches.

Post hoc analyses revealed that shadow accounts for 62.4 percent and projection for 18.1 percent of the total variation in accuracy. Additionally, the combined total of the higher order interactions accounts for 19.4 percent of total variation in accuracy. From the rankings we see that shadows are the most important factor, followed by perspective projection.

Experiment 2: Rotation

We calculated the angular difference between the normal of a face of the fixed cube and the normal of the corresponding face on the test cube as a measure of orientational accuracy. We ran a repeated measures MANOVA on the measure, with the pictorial cues serving as repeating factors. We found three main effects: shadow, projection, and motion (see Table 4). Elevation, object texture, and ground texture did not have statistically significant effects (all had $p > 0.4$). We also found some higher order interactions (see Table 5).

We ran post hoc analyses to determine the directions of the main effects. Matched pairs *t* tests revealed that the presence of shadows reduced the angular difference, but that the reduction was not statistically significant ($t = 0.861$, $df = 766$, $p = 0.389$). Masking is the most likely explanation for shadow being a significant factor in the MANOVA and not in the *t* test. Because the *t* test does not take into account the effects of other dependent variables, stronger factors can "mask" the effect of weaker factors.

Table 1. Positioning task main effects.

Factor	$F(1,11)$	p
Shadow	112.552	0.0001
Projection	21.533	0.0007

Table 2. Positioning task higher order interactions.

Factor	$F(1,11)$	p
Object texture by projection	6.587	0.0262
Shadow by projection by motion	6.804	0.0243
Object texture by ground texture by shadow by projection	9.998	0.0090

Table 3. Positioning task mean accuracies.

Factor	Mean Error (in inches)	Std. Dev. (in inches)
Without shadows	0.546	0.351
With shadows	0.299	0.216
Orthographic projection	0.494	0.351
Perspective projection	0.351	0.299

We believe the MANOVA is the more reliable test in this case.

Further matched pairs *t* tests determined that the presence of motion significantly reduced the angular difference ($t = 2.368$, $df = 766$, $p < 0.018$) and that the presence of perspective projection significantly increased the angular difference ($t = 7.521$, $df = 766$, $p < 0.001$). Table 6 shows the mean accuracies for the main effects.

Additional post hoc analyses of the total variation in accuracy revealed that shadow accounts for 0.8 percent, projection for 55.3 percent, and motion for 5.9 percent. Additionally, higher order effects account for 36.3 percent of the total variation in accuracy.

As in Experiment 1, the randomly varying object position warrants an ANOVA to separate the effects of the randomly varying position from the spatial cues (see above). The *y* and *z* axis rotations did not have significant effects (both had $p > 0.3$).

Table 4. Rotation task main effects.

Factor	$F(1,11)$	p
Projection	9.493	0.0116
Motion	14.824	0.0032
Shadow	5.176	0.0462

Table 5. Rotation task higher order interactions.

Factor	$F(1,11)$	p
Ground texture by motion	10.666	0.0085
Projection by motion	8.802	0.0141
Object texture by shadow by projection	5.148	0.0467
Object texture by ground texture by motion	14.333	0.0036
Elevation by object texture by ground texture by projection	6.468	0.0292

Table 6. Rotation task mean accuracies.

Factor	Mean Error (in degrees)	Std. Dev. (in degrees)
Orthographic projection	4.811	3.078
Perspective projection	6.574	3.409
Without motion	5.979	3.430
With motion	5.406	3.274
Without shadows	5.797	3.432

Experiment 3: Scaling

We used the difference between the projected area of the fixed ball and the projected area of the test ball as a measure of sizing accuracy. We ran a repeated measures MANOVA on the measures, with the pictorial cues used as repeating factors. We found five main effects: elevation, object texture, shadow, projection, and motion (see Table 7). Ground texture did not have statistically significant effects ($p > .05$). Additionally, there were a large number of higher order interactions (see Table 8).

We ran post hoc analyses to determine the directions of the main effects. Matched pairs *t* tests revealed that the presence of object texture, shadow, and motion significantly decreased the difference of areas. Specifically, $t = 3.019$ and $p = 0.051$ for object texture, $t = 2.737$ and $p = 0.006$ for shadow, $t = 1.952$ and $p = 0.051$ for motion, and $df = 766$ for all three cues. On the other hand, elevation and perspective projection significantly increased the difference of areas ($t = 3.640$ and 4.442, respectively, $df = 766$, and both $p < 0.001$). Table 9 shows the mean accuracies for the main effects.

Further post hoc analyses of the total variation in accuracy revealed that elevation accounted for 5.3 percent, object texture for 2.2 percent, shadow for 2.9 percent, projection for 7.5 percent, and motion for 0.9 percent. Additionally, higher order interactions accounted for 81.2 percent of the total variation in accuracy. The presence of so many main and crossed effects and the fact that crossed effects accounted for such a large percentage of the total variation in accuracy indicates the importance of all the cues. It also indicates that the cues influenced each other's salience.

As in Experiments 1 and 2, the randomly varying object position warranted ANOVA to separate the effects of the

Table 7. Scaling task main effects.

Factor	$F(1,11)$	p
Object texture	6.107	0.0311
Shadow	6.990	0.0228
Motion	7.716	0.0180
Elevation	7.999	0.0164
Projection	18.160	0.0013

Table 8. Scaling task higher order interactions.

Factor	$F(1,11)$	p
Object texture by ground texture	9.463	0.0105
Object texture by shadow	10.283	0.0084
Object texture by projection	9.296	0.0111
Ground texture by projection	12.039	0.0052
Shadow by projection	10.178	0.0086
Object texture by motion	9.730	0.0098
Projection by motion	6.020	0.0320
Elevation by object texture by shadows	7.552	0.0190
Elevation by ground texture by projection	13.852	0.0034
Ground texture by shadow by motion	12.217	0.0050
Elevation by projection by motion	8.866	0.0126
Ground texture by projection by motion	18.328	0.0013
Elevation by object texture by ground texture by projection	16.308	0.0020
Elevation by object texture by ground texture by motion	7.356	0.0202
Elevation by ground texture by shadow by motion	13.745	0.0035
Elevation by shadow by projection by motion	16.293	0.0020
Elevation by object texture by ground texture by shadow by projection by motion	7.831	0.0173

randomly varying position from the spatial cues. The *y* and *z* axes rotations had no significant effects (both had $p > 0.7$).

Reference

1. R.B. Darlington, *Regression and Linear Models*, McGraw Hill Series in Psychology, McGraw Hill, New York, 1990.

Table 9. Scaling task mean accuracies.

Factor	Mean Error (in sq. inches)	Std. Dev. (in sq. inches)
Without object texture	0.090	0.090
With object texture	0.087	0.096
Without shadows	0.108	0.106
With shadows	0.068	0.072
Without motion	0.096	0.099
With motion	0.081	0.086
On the ground	0.083	0.085
In the air	0.094	0.100
Orthographic projection	0.087	0.092