FISH TANK VIRTUAL REALITY

Colin Ware
Faculty of Computer Science
P.O Box 4400
University of New Brunswick
Canada E3B 5A3
email: cware@unb.ca

Kevin Arthur and Kellogg S. Booth
Department of Computer Science
6356 Agricultural Road
University of British Columbia
Vancouver, B.C.
Canada V6T 1Z2
email: karthur@cs.ubc.ca
ksbooth@cs.ubc.ca

ABSTRACT

The defining characteristics of what we call "Fish Tank Virtual Reality" are a stereo image of a three dimensional (3D) scene viewed on a monitor using a perspective projection coupled to the head position of the observer. We discuss some of the relative merits of this mode of viewing as compared to head mounted stereo displays. In addition, we report the experimental investigation of the following variables: 1) whether or not the perspective view is coupled to the actual viewpoint of the observer, 2) whether stereopsis is employed. Experiment 1 involved the subjective comparison of pairs of viewing conditions and the results suggest that head coupling may be more important than stereo in yielding a strong impression of three dimensionality. Experiment 2 involved subjects tracing a path from a leaf of a 3D tree to the correct root (there were two trees intermeshed). The error rates ranged from 22% in the pictorial display, to 1.3% in the head coupled stereo display. The error rates for stereo alone and head coupling alone were 14.7% and 3.2% respectively. We conclude that head coupling is probably more important than stereo in 3D visualization and that head coupling and stereo combined provide an important enhancement to monitor based computer graphics.

KEYWORDS: virtual reality, scientific visualization, head coupled displays, stereopsis.

INTRODUCTION

J.J. Gibson's pioneering research showed the interrelatedness of perceptual systems. Information from a variety of systems, including the kinaesthetic feedback relating to self directed body movement coupled to image changes are crucial to our understanding of space (Gibson, 1979). Recently, Deering (1992) presented the technical components required to create a high quality 3D "virtual reality" image on a monitor by tracking the head of the user. He emphasized the importance of the accurate coupling of the perspective image to the observer's viewpoint. We have been using the term "Fish Tank VR" to describe the same kind of display - one in which the virtual 3D scene is obtained by coupling head position with respect to a monitor to the 3D image displayed so that the correct perspective view is obtained (see also McKenna, 1992). The resulting scene can be either viewed monocularly, coupled to a single eye position, or binocularly, if suitable stereo equipment is available. Figure 1 shows the basic setup, which includes a monitor, a device for measuring head position, and stereo viewing goggles. Fish Tank VR has a number of advantages over immersion VR. We begin by describing these advantages to show why we should be interested in the properties and uses of this mode of viewing.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

© 1993 ACM 0-89791-575-5/93/0004/0037...$1.50

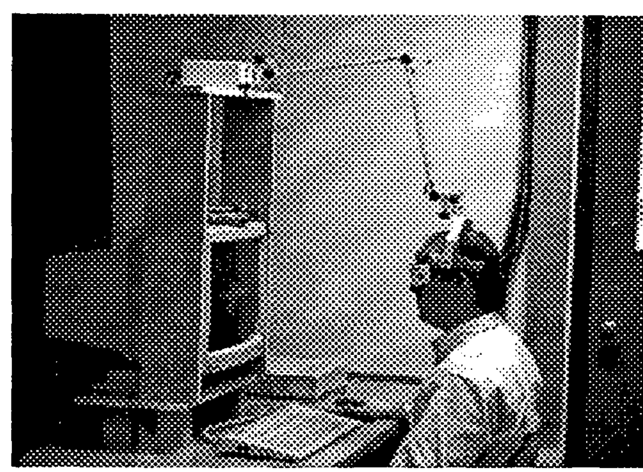

Figure 1. The head coupled display system. The subject's head position is monitored by the ADL-1. StereoGraphics glassed provide the stereo when used with a monitor (and graphics system) capable of a 120 Hz update rate, 60 Hz to each eye.

Resolution

In immersion VR with a head mounted display, the monitors are placed very close to the eyes giving a field of view which may subtend 90 deg of visual angle (Ward et al, 1992, Sutherland, 1968). Given the typical resolution of the current displays each pixel will subtend approximately 12 minutes of arc. Viewing a high resolution monitor with a 30 deg field of view yields 2 minutes of arc per pixel - close to the resolution limits of the human eye. While it is true that the resolution of head mounted displays for immersion VR will improve over time, it seems likely to be a long time before it can be expected to come close to current technology Fish Tank VR.

Depth-of-Field

Depth-of-field effects arise from the fact that we can change the focal length of our eyes and therefore things we are fixating will typically be in focus, while things we are not fixating will be out of focus. Thus if nearby objects are fixated, background objects should be out of focus. Without directly measuring the focal length of the observer's lens, it is impossible to take depth-of-field directly into account in creating images for VR. However, because of the viewing geometry in Fish Tank VR the working scene is necessarily constrained to lie within a few centimeters in front and behind the screen of the monitor, this is because things that are nearer and further away are clipped when the subject makes a head movement. It is possible to simulate depth of field effect by drawing a background that approximates an out-of-focus image.

Stability in the Presence of Eye Movements

In immersion VR the eyes are necessarily very close to the display monitors (which are mounted on a helmet) in order to get the wide field of view. This introduces an error when the eye moves off axis. As Deering (1992) has pointed out, the eye rotates about a geometric center, which is about 6.0 mm behind the optic center (first nodal point). This means that a 40 deg eye movement will result in a 3.8 mm translation of the center of the lens. Coupled with a screen placed 8 cm away from the eye this will result in a position change of nearly 3.8 cm for an object at 80 cm. This effect can only be corrected by directly measuring eye movements (which can be done but which adds considerable complication and expense). In Fish Tank VR the same effect exists, but is much smaller, only amounting to a 3.8 mm movement for objects at 80cm if they are close to the plane of the screen.

Integration of the VR Workspace with the Everyday Workspace

Immersion VR has the major advantage of a wide field-of-view which can give the feeling of existing in the graphical world. However, the cost of this is to block out the everyday world of desks, chairs and filing cabinets with the consequence that the inhabitants of VR have to have handlers to make sure that they do not hurt themselves. The Fish Tank VR workspace can be part of the office, just as the workstation can be part of the office. (We note that work is also progressing on "augemented VR", the purpose of which is to blend everyday reality and graphics using head mounted partially transparent displays with a wide viewing angle.)

DESIGN PROPERTIES OF OUR FISH TANK VR

Our first experiment was designed to investigate two of the factors leading to the subjective impression of three dimensional space, namely the relative importance of stereopsis and head coupling. We wished to examine these factors in the context of a scene that would seem as convincingly 3D as possible, that is, it should have most of the factors that contribute to our impression of space, such as appropriate shading, shadows, and other spatial cues in addition to head coupling and stereo. This is unlike the normal study which starts with an impoverished scene and adds only a single factor, such as stereopsis. Instead, our study can be regarded as taking a scene which is rich in spatial and temporal information and subtracting either stereopsis, head coupled perspective, or both.

In the following sections we present a brief discussion of the various design decisions which went into constructing our Fish Tank VR experiments.

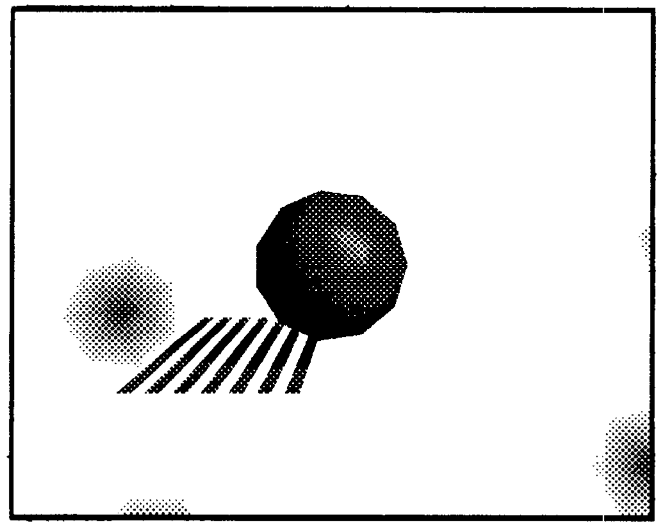

Figure 2. The sphere display used in the first experiment. Hardware lighting was used to achieve the specular reflection, while the fuzzy cast shadow was pre-computed.

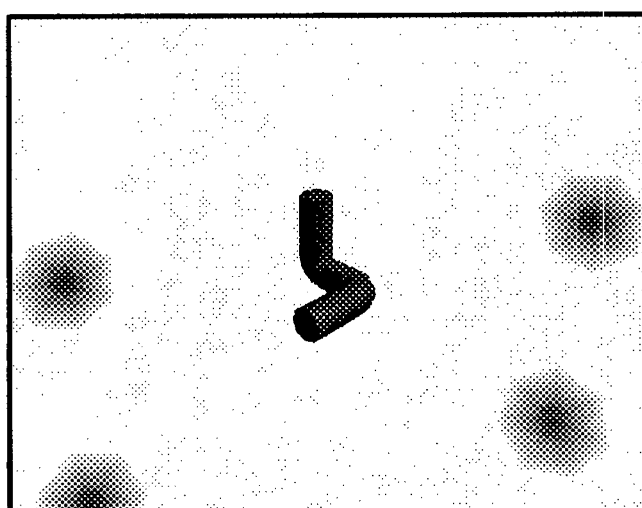

Figure 3. The Shepard-Metzler bent tube display used in the first experiment.

Objects

We constructed the two very simple scenes shown in Figures 2 and 3 for our first experiment: one consisted of a sphere with its shadow cast on a set of parallel horizontal bars below and to the left; the other consisted of a bent piece of tube based on Shepard and Metzler's (1971) mental rotation objects. Both of these scenes were given what we call our "Vection Background". This requires a little explanation.

The term "vection" is usually associated with the feeling of self movement when a large field display is moved with respect to an observer. Thus people placed in the center of a drum which rotates independent of them will, under the right circumstances, feel that it is they who are rotating, not the drum. It is often claimed that it is the wide-field-of-view peripheral stimulus that is critical for the vection experience. However, recent evidence suggests that the effect can be achieved with a small field-of-view (Howard and Heckman, 1989). Also, it is commonly observed that the vection effect can be obtained when looking out of a small airplane window across the aisle when another plane starts to move.

Howard and Heckman suggested that one of the important factors in eliciting vection is the perceived distance of a moving visual image. Images that are perceived as furthest away contributing the most. In the case of Fish Tank VR, we wish the observer to feel that the monitor is a window to an extensive space, so we reasoned that the experience of spaciousness would be enhanced with an appropriate background. Accordingly, we generated a background that consists of a random field of objects computed as though they were at infinity with respect to the observer. We call this our "vection" background.

Depth-of-Field

As mentioned in the introduction, one of the advantages of Fish Tank VR is that it constrains the field of interest to the region within the frame of the monitor, and a few centimeters in front and behind it. With this assumption, more distant objects will be out of focus. Accordingly, we made the vection background out of fuzzy discs, to give the illusion of depth-of-field. It should be understood that the vection background is not intended to be focussed on; instead it is intended to give a feeling of spaciousness when objects in the foreground are fixated.

Shading

Three types of shading information seem to be important in aiding our perception of the layout and shape of objects in space. Shading, both Lambertian and specular, tells us about the shape of surfaces, while cast shadows tell us about the positions of objects relative to each other (Wanger et al., 1992).

Spatio-Temporal Accuracy in Head Tracking

Deering (1992) presents a strong case for accuracy in both time and space being important in obtaining a strong percept of 3D space. If there are distortions in the measured position of the head - and hence the viewer's eye, then the result will be a scene that appears to be made of rubber and which flexes as the viewer moves. Performance decrements can be expected if there is temporal lag in the device that measures head position (Smith, 1962), although these may possibly be mitigated by the use of a predictive filter (Liang et al., 1991; Friedman, et al., 1992) One of the most common devices used to track head position is the 3Space Isotrack, a six degree of freedom position sensor made by Polhemus. However, it is known that this device gives position information with a significant temporal lag, which may be as much as 80 msec behind the current position (Liang et al., 1991). We used the Shooting Star Technology ADL-1 to provide head position information without any software smoothing. This device uses mechanical linkages with potentiometers at the joints to give fast readings (see Figure 1). The lag in this device is small (2ms), much less than that induced by other factors such as the time taken to read the input buffer and to update the image on the screen. Its rated absolute positional accuracy is 0.51 cm and its resolution is 0.064 cm.

Stereo Display

To obtain a stereo view we used the StereoGraphics CristalEyes stereo system coupled with the SGI workstation's ability to display 120 frames/second (60 to each eye). Unfortunately this mode of stereo display is far from perfect and some significant ghosting can be seen from the image designated for the left eye bleeding into the right eye view, and vice versa. To minimize this effect we chose colours with only a small green component since the green phosphor has the longest decay rate on our monitor (and this is typical of most monitors - a better solution would have been to replace the monitor with one more suited for stereo display, as recommended by Deering (1992)).

High Frame Rate

Because of the desire of users of 3D stereo display equipment to display complex information, they almost always show so much information that the frame update rate drops below the maximum of 60 Hz. In fact, frame rates of only 10 Hz are more typical.

One of our design criterion was to create a scene which was reasonably rich in 3D spatial cues, but which still had a fast update rate. By carefully limiting the complexity of the scenes we were able to achieve the 60 Hz update rate with the scenes illustrated using a Silicon Graphics 4D 240 VGX.

EXPERIMENT 1: SUBJECTIVE IMPRESSIONS OF THREE DIMENSIONALITY

To compare the relative effectiveness of head coupling and stereopsis we designed an experimental protocol allowing subjects to make comparisons between pairs of presentation methods, toggling between them until they decided which contributed more to the perception of 3D space. We also questioned subjects after the experiment on their feelings about the value of the different modes of display.

The five conditions are given in Table 1. The experiment was carried out with the subjects always wearing the stereo goggles and the head tracking system. In the non-stereo conditions the same scene was presented to the two eyes. In the binocular non-stereo viewing condition the viewpoint was between the eyes. In the monocular viewing condition, the viewpoint was correct for the right eye and the subjects was asked to "Close your left eye". In the fixed viewpoint condition the perspective view was established by the subject's head position at the start of the trial. Subjects were asked to move their heads around for all conditions in order to assess the value of head coupling.

Table 1: Experimental Conditions

1	Picture
2	Stereo only
3	Head Coupled Monocular
4	Head Coupled Binocular
5	Head Coupled + Stereo

Trials

On a given trial subjects were allowed to toggle between two viewing conditions using the space bar, continuing to examine each until they had decided which gave the strongest impression of three dimensionality. When they had decided which was "best", they made the selection using the mouse. This automatically advanced them to the next pair of conditions. There were 10 pairwise comparisons of 5 conditions. A trial block was made up 20 trials consisting of the 10 pairwise comparisons for the sphere scene and the 10 pairwise comparisons for the bent tube scene. The entire block of 20 trials was repeated twice. The order of all comparisons was randomized.

Following the comparison trials, subjects were asked the following set of questions and their answers were recorded.

All of the following questions relate to the quality of the 3D spatial impression

Is head coupling as important, more important or less important than stereo?

Is the combination of head coupling and stereo better than either alone?

Is head coupling alone worthwhile? (If you had the option would you use it?)

Is stereo alone worthwhile? (If you had the option would you use it?)

Is head couplng with stereo worthwhile? (If you had the option would you use it?)

Do you have other comments on these methods of displaying 3D data?

Seven subjects were used in this experiment, four of whom were well acquainted with high performance graphics systems

RESULTS FROM EXPERIMENT 1

There were no systematic differences between the results obtained from the two scenes and so these data have been merged. The results are summarized in Table 2, which represents the combined data from all subjects. This matrix shows for each pair of conditions which condition gave the strongest 3D impression. Thus the value 89% in row 4 column 2 means that in 25 out of 28 possible responses subjects found the Head Coupled (non-stereo) display more compelling than the Stereo only (non-head coupled) display.

Table 2

All Subjects	1	2	3	4	5	All
1 Picture		43%	1%	0%	7%	13%
2 Stereo only	57%		7%	11%	0%	19%
3HC monocular	96%	93%		29%	61%	70%
4 HC binocular	100%	89%	71%		68%	82%
5 HC + stereo	93%	100%	39%	32%		66%

What is most striking about this data is the fact that static stereo was rarely found to be more effective than head coupled viewing without stereo. The right hand column shows for each condition the overall percentage of responses for which that condition was preferred. This shows that head coupled displays without stereo were preferred somewhat more often than head coupled displays with stereo. This may possibly be attributed to the ghosting of the image which occurs due to imperfect phosphor decay, causing cross talk between the left and right eye images.

The results from the set of questions also strongly supported the usefulness of head coupled stereo viewing. All users said that they would use it for object visualization if it were available. When asked to compare the importance of head coupling with stereo, two of the seven subjects stated that they thought stereo was more important than head coupling. However, these same subjects preferred the head coupling in the direct comparison task. One subject complained about the awkwardness of the apparatus and pointed out that that would be a factor in how often it would be used.

EXPERIMENT 2: TRACING TREE PATHS

For our second experiment we chose a task that was designed by Sollenberger and Milgram (1992) to study the ability of observers to perceive arterial branching in brain scan data under different viewing conditions. The task involves the construction of two trees in 3D space, whose branches overlap considerably. A leaf of one of the trees is marked and the subject has to determine to which of the two tree roots that branch belongs. Errors are measured to assess the conditions. Sollenberger and Milgram used this task to look at trees viewed with and without stereo and with and without rotation. We used it to look at head coupled perspective viewing and stereo using the same set of conditions used in experiment 1. In addition to errors, we measured task performance time.

Tree construction: Our trees were recursively defined ternary trees. Each parent node had three child nodes connected to it by lines. The height of each child above its parent is 70% of the height of the parent above the grandparent. The lateral positions of the children were randomly placed relative to the parent. There were three levels of branches above the trunk, resulting in 27 leaves for each tree.

The following recurrence relation gives a precise specification.

$\mathbf{HorizontalSpacing}_{root} = 8.0$ cm.
$\mathbf{VerticalSpacing}_{root} = 8.0$ cm.

$\mathbf{HorizontalSpacing}_{child} = 0.7 * \mathbf{HorizontalSpacing}_{parent}$
$\mathbf{VerticalSpacing}_{child} = 0.7 * \mathbf{VerticalSpacing}_{parent}$

$Y_{child} = Y_{parent} + \mathbf{VerticalSpacing}(1.0 + \text{Rand}()*0.25)$
$X_{child} = X_{parent} + \mathbf{HorizontalSpacing}*\text{Rand}()$
$Z_{child} = Z_{parent} + \mathbf{HorizontalSpacing}*\text{Rand}()$

where Rand() returns a uniform random number in the range [-1,+1]

In addition a vertical trunk of length **VerticalSpacing** is added at the base of the tree.

On each trial two trees were constructed with roots laterally separated by one cm. A yellow circle was placed on the leaf closest to the midline between the two roots, only taking the x coordinate into account. The reason for this was to eliminate trials that would be easy in all conditions because they occurred in parts of the trees where there was no overlap with the other tree. An example of a pair of trees photographed from the monitor screen is shown in Figure 4. The trees were coloured purple (monitor red plus monitor blue) on the same background as used in Experiment 1. The triangle and square at the tree roots were coloured green.

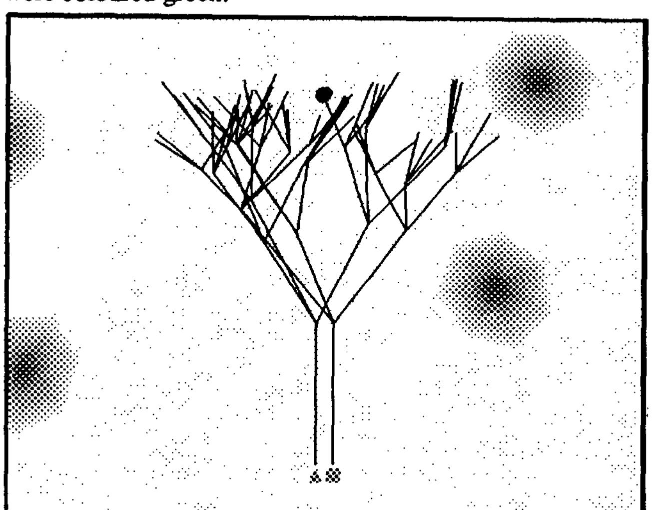

Figure 4. An example of a tree display used in Experiment 2. The purple colour was used to minimize ghosting.

The five viewing conditions listed in Table 1 were employed. Ten subjects who consisted of graduate and undergraduate students were instructed to be as accurate as they could and not to worry about how long they were taking.

Trial Blocks

A practice group of 10 trials (two in each condition) was given at the start of the experiment. Trials were given in groups of 22, with the first two trials of each group designated as additional practice trials and where all 22 trials were given in one of the five viewing conditions. A trial block consisted of all 5 groups given in a random order, and the entire experiment consisted of 3 such blocks resulting in a total of 60 trials in each of the 5 experimental conditions.

RESULTS FROM EXPERIMENT 2

The results from Experiment 2 are summarized in Tables 3 and 4. The timing data shows that the head coupled stereo condition was the fastest, but that head coupling alone was slow. There are significant differences at the 0.05 level between both conditions 3 and 4 and condition 5 (by the Wilcoxon matched pairs signed ranks test). The only other difference that is significant is between condition 4 and condition 1.

Table 3

Timing data	times (sec)
1 Picture	7.50
2 Stereo only	8.09
3 HC monocular	8.66
4 HC binocular	9.12
5 HC + stereo	6.83

Table 4

Error data	% errors
1 Picture	21.8
2 Stereo only	14.7
3 HC monocular	3.7
4 HC binocular	2.7
5 HC + stereo	1.3

The error data is more interesting, with errors ranging from 21.8% in the static, no stereo condition, to 1.3% for the head coupled stereo conditions. All of the differences are significant in pairwise comparisons except for the difference between conditions 3 and 4, the two head coupled conditions without stereo.

Overall, the error rates obtained are lower than those obtained by Sollenberger and Milgram (1991), but the pattern is strikingly similar despite the differences in the stimulus trees, the viewing condition and the experimental protocols. There are two other similarities between our findings and those reported by Sollenberger and Milgram: we found motion to be more important than stereo, even though their motion was simple rotation of the object whereas ours resulted from head coupling; and we found the combination of motion (head coupling in our case) and stereo to be more effective than either in isolation.

DISCUSSION

The strong preference expressed by most subjects for the head coupled displays over the stereo displays in Experiment 1 and the enthusiastic response of viewers to a head coupled display both suggest that people who observe 3D scenes with graphics systems capable of real-time update rates should consider investing in some method of tracking head position and in coupling the displayed image directly to the viewpoint of the observer. Once this is done the subjective results suggest that stereopsis may add only marginally to the perception of three dimensionality of objects.

Experiment 2 provides objective evidence that head coupled stereo can help users to comprehend a complex tree structured graphical object. Here the evidence shows that both head coupling and stereopsis contribute to performance. The task maps well into two domains of considerable current interest, the domain of medical imaging where doctors may wish to trace blood vessels in brain scan data, and the domain of 3D software visualization where software engineers may wish to trace object dependencies between software modules represented as networks in 3D space (Robertson, et al., 1991, Fairchild et al., 1988). In both applications error rates are critically important and our finding is that head coupled stereo can reduce error rates by a factor of sixteen over a static pictorial display.

It can be argued that it is the motion-induced depth (Wallach and O'Connell, 1953) and not the head coupling as such that produced both the improved spatial percept and the improved performance on the tracing task. Our current evidence does not counter this objection. However, it is likely the head coupled image motion (note that the object appears fixed in space, it is the image which moves) is a way of providing spatial information which is more appealing than displaying the scene rocking back and fourth about a vertical axis, as is commonly done in molecular modelling packages.

ACKNOWLEDGEMENTS
Financial support for this project was provided by the Natural Sciences and Engineering Research Council of Canada and the British Columbia Advanced Systems Institute.

REFERENCES

1. Deering, M. (1992) High Resolution Virtual Reality. Computer Graphics, 26, 2, 195-202.

2. Fairchild, K.M., Poltrock, E.E. and Furnas, G.W. (1988) SemNet: Three-Dimensional Graphic representations of large knowledge bases. In Cognitive Science and Its Applications for Human-Computer Interaction. Ed Raymond Guindon, Lawrence Erelbaum, 201-233.

3. Friedman, M., Starner, T and Pentland, A., (1992) Device Synchronization using and optimal linear filter. 1992 Symposium on Interactive 3D graphics, Special Issue of Computer Graphics, 57-62.

4. Gibson, J.J. (1979) The Ecological Approach to Visual Perception. Houghton Mifflin: Boston.

5. Howard, I.P. an Heckman, T. (1989) Circular vection as a function of the relative sizes, distances and positions of two competing visual displays. Perception, 18 (5) 657-665.

6. Liang, J., Shaw, C., and Green, M. (1991) On temporal realism in the virtual reality environment. UIST'91 19-25.

7. McKenna, M, (1992) Interactive viewpoint control and three-dimensional operations. 1992 Symposium on Interactive 3D graphics, Special Issue of Computer Graphics, 53-56.

8. Robertson, G.G. Mackinlay, J.D. and Card, S.K. (1991) Cone trees: Animated 3D visualizations of hierarchical information. Proc of CHI'91: 189-194.

9. Shepard, R.N. and Metzler, J. (1971) Mental rotation of three-dimensional objects. Science, 171, 701-703.

10. Smith, K.U. (1962) Delayed sensory feedback and behavior, W.B. Saunders, Philadelphia.

11. Sollenberger, R.L., and Milgram, P. (1991) A comparative study of rotational and stereoscopic computer graphics depth cues. Proceedings of the Human Factors Society 35th Annual Meeting, San Fancisco, 1452-1456.

12. Sutherland, I. (1968) A head-mounted three dimensional display. Fall Joint Computer Conference, AFIPS Conference Proceedings, 33, 757-764.

13. Wallach, H. and O'Connell, D.H. (1953) The kinetic depth effect. Journal of Experimental Psychology, 45, 205-217.

14. Wanger, L.R. Fewerda, J.A. and Greenberg, D.P. (1992) Perceiving spatial relationships in computer generated images. IEEE Computer Graphics and Applications, 12(3) 44-59.

15. Ward, M., Azuma, R., Bennett, R., Gottschalk, S., and Fuchs, H. (1992) A demonstrated optical tracker with scalable work area for head mounted display systems. 1992 Symposium on Interactive 3D Graphics, Special Issue of Computer Graphics, 43-52.