The "Silk Cursor": Investigating Transparency for 3D Target Acquisition

Shumin Zhai[1] *William Buxton*[2] *Paul Milgram*[1]*

[1]Department of Industrial Engineering
University of Toronto
Toronto, Ontario, M5S 1A4, Canada
Tel: 1-416-978-3776
E-mail: zhai@ie.toronto.edu
milgram@ie.utoronto.ca

[2]University of Toronto & Xerox PARC
c/o CSRI
University of Toronto
Toronto, Ontario, M5S 1A4, Canada
Tel: 1-416-978-1961
E-mail: buxton@dgp.toronto.edu

ABSTRACT
This study investigates dynamic 3D target acquisition. The focus is on the relative effect of specific perceptual cues. A novel technique is introduced and we report on an experiment that evaluates its effectiveness.

There are two aspects to the new technique. First, in contrast to normal practice, the tracking symbol is a volume rather than a point. Second, the surface of this volume is semi-transparent, thereby affording occlusion cues during target acquisition.

The experiment shows that the volume/occlusion cues were effective in both monocular and stereoscopic conditions. For some tasks where stereoscopic presentation is unavailable or infeasible, the new technique offers an effective alternative.

KEYWORDS: 3D interface, interaction technique, target acquisition, virtual reality, Fitts' law, input, depth perception.

INTRODUCTION

With the advent of modern workstations and increasing demands for computer-based applications, 3D techniques are moving from the restricted domain of graphics to mainstream applications (e.g.[2]). However, as we move to 3D, we see a breakdown in many of the interaction techniques that have traditionally been used in 2D direct manipulation systems. Tasks such as inking, target acquisition, pursuit tracking, sweeping out regions, orientation, navigation and docking present new challenges to the interaction designer.

Largely as an outgrowth of computer graphics, a body of research is developing which is beginning to address some of the interaction issues confronting the designer. Representative examples are found in [4, 8, 11, 13]. However, there remain large gaps both in the literature and in practice, and of the techniques described, there has been little in the way of experimental evaluation. Chen, Mountford & Sellen [3] is one of a few notable exceptions.

In the study reported below, our intent is to contribute to this body of research. We introduce a new technique for dynamic 3D target acquisition. After describing the technique, we report on an experiment that evaluates its effectiveness under various circumstances.

The technique described is novel in two respects. First, the tracking symbol used is a *volume* rather than a point, as is the case in conventional systems. Second, the surface of the tracking volume is *semi-transparent*, thereby providing additional depth cues beyond what is achievable with conventional techniques, primarily due to partial occlusion of the tracking volume by the object being tracked.

After a brief discussion of these two novel aspects of our design, we proceed to describe the experiment in which the technique which uses them was tested.

The "Prince" Technique: the Cursor as Region or Volume

One of the most studied aspects of HCI is target acquisition using Fitts' Law [5]. According to this law, the movement time between two targets of width "W," separated by amplitude "A", can be modelled as follows [7]:

$$MT = a + b \log_2(A / W + 1)$$

This is illustrated in Fig. 1(a).

In this paper, we introduce a variation on the conditions to which this model pertains. As is illustrated in Fig. 1(b), we have reversed the situation such that the objects being selected are *points* (separated by distance "A"), and the cursor is a *region* of width "W".

*Currently on leave at ATR Communication System Research Laboratories, Kyoto, Japan

Our assumption is that Fitts' Law holds under this new condition. We have dubbed this the "Prince" technique, after the first company to make over-size tennis rackets (which epitomize the underlying principle).

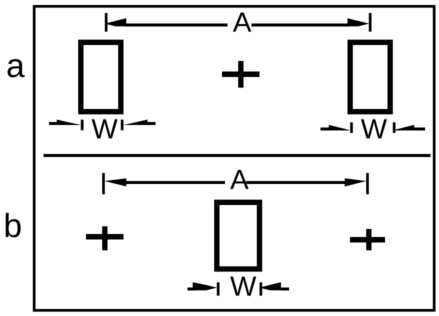

Figure 1. Two representations of Fitts' Law. The top half (a) shows the traditional representation. Targets of width "W" are selected by the cursor (the point defined by the "+"), across amplitude "A". In the lower half (b), two points (represented by the two "+" symbols), separated by amplitude "A" are selected by a cursor of width "W".

We believe the idea of using a region for the cursor has value in cases where one is selecting individual points, or small objects, or collections of points. We also believe that the concept will extend to 3D by having the cursor be a *volume*.

The "Silk Stocking" Effect: Using Occlusion for Enhanced Depth Cues

The second novel aspect of the technique which we introduce is our proposed means for obtaining depth cues through the use of occlusion. The simplest way to describe our technique is to imagine that the surface of the cursor's volume is covered by a nearly transparent material, like a silk stocking. This is illustrated in Fig. 2.

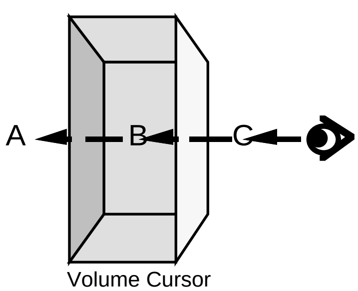

Figure 2: Using a "silk" covering over a rectangular volume cursor in order to obtain occlusion-based depth cues. An object at point A is seen through two layers of "silk", and so is perceived to be *behind* the volume cursor. An object at point B is seen through one layer, and so is perceived as *inside* the cursor's volume. An object at point C is not occluded by the silk at all, and so is seen to be *in front* of the volume.

Using this technique, which has recently been enabled by the power of modern graphics workstations, one can easily tell if an object is behind, inside or in front of the cursor. (see colour plates).

EXPERIMENT

Experimental Hypothesis

The primary goal of our experiment was to evaluate the effectiveness of the silk surface on a volume cursor in a 3D dynamic target acquisition task. We tested the volume cursor both with the silk surface and without it (i.e., in an outline "wire frame" version). Since stereoscopic projection is widely recognised as one of the most effective and common 3D interface techniques[12, 14, 15], we tested each in both mono and stereo display conditions. Thus, the experiment had four conditions: stereo display with silk cursor (SS), stereo display with wire frame cursor (SW), mono display with silk cursor (MS) and mono display with wire frame cursor (MW).

Our hypothesis was that the silk-like surface and the stereoscopic display would each significantly improve 3D target acquisition, and that the two factors together would enhance each other. What was of particular interest to us was whether or not the silk surface effect alone (i.e. the MS condition) would generate superior, or in any case comparable, performance to the SW condition, which would confirm to us the potential advantages of the silk cursor on its own as a 3D target acquisition technique.

Experimental Task

A 3D dynamic target acquisition task, "virtual fishing", was designed for the experiment. In each trial of the experiment, an “angel fish” with random size and color appears swimming around randomly within a 3D virtual environment, as shown in Fig. 3.

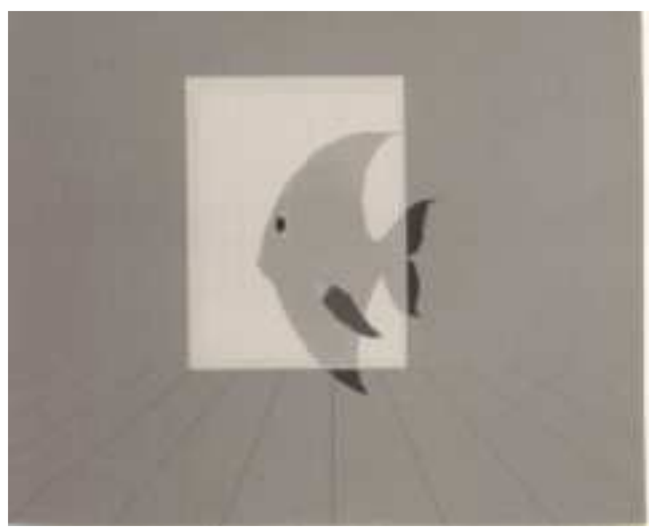

Figure 3: The "fishing" task

The subjects were asked to move a 3D volume cursor to envelop the fish and "grasp it" when the fish was perceived to be completely inside of the cursor. Subjects wore a special glove, and "grasping" was done by closing the hand.

If the fish was entirely inside of the cursor volume, the trial was successful and the fish stayed "caught" inside of the cursor. The time score of the trials was displayed to the subject, along with a short beep. If the fish was not completely inside of the cursor when grasped, the fish disappeared. In this case, which was considered a "miss", a long beep was sounded and error magnitude in each x, y, and z dimension was displayed. Subjects pressed the spacebar on the workstation to activate each new trial.

The origin of the {x y z} coordinate system was located at the center of the computer screen surface, with positive x axis pointing to right, y pointing up and z pointing to the user. All objects were drawn using polar projection and were modelled in units of centimeters, where 1 cm in the virtual fish tank corresponded to 1 cm in the real world for any line segment appearing within the same plane as the surface of the screen. The x (from lips to tail end), y (vertical) and z (from left fin tip to right fin tip) dimensions of the largest “adult” fish were 10 cm, 15 cm and 1.3 cm respectively. The smallest “baby” fish was 30 percent of the size of the largest “adult” fish. The cursor had a constant size of 11.3 cm, 16.3 cm and 2.6 cm in x, y and z dimensions.

The fish movement was driven by independent forcing functions in the x , y and z dimension. Since a suitable combination of sine functions generates smooth subjectively unpredictable motion, it has been conventionally used in manual tracking research [9]. in this experiment, forcing functions applied to the fish motion were:

$$x(t) = \sum_{i=0}^{5} Ap^{-i} \sin(2\pi f_0 p^i t + \phi_x(i))$$

$$y(t) = \sum_{i=0}^{5} Ap^{-i} \sin(2\pi f_0 p^i t + \phi_y(i))$$

$$z(t) = -7.8 + \sum_{i=0}^{5} Ap^{-i} \sin(2\pi f_0 p^i t + \phi_z(i))$$

where t is time. A = 4.55 cm, p = 2, f_o = 0.02 Hz; $\phi_x(i)$ $\phi_y(i)$ and $\phi_z(i)$ are pseudo-random numbers, ranging uniformly between 0 and 2π. (For a more detailed explanation of these forcing functions, see [17]).

In silk cursor mode, the semi-transparent surface intensity I was rendered by interpolating the cursor colour intensity (source) I_s with the destination colour intensity I_d [6] according to:

$$I = \alpha I_s + (1 - \alpha) I_d$$

The α coefficient was set at 0.38 for all surfaces of the cursor, except the back surface for which α was set at 0.6.

Experiment Platform

The experiment was conducted using the MITS (Manipulation in Three Space) system [16, 17] developed by the authors. MITS is a non-immersive stereoscopic virtual environment, based on a SGI IRIS 4D Crimson/VGX graphics workstation equipped with CrystalEyes™ stereoscopic glasses. In this experiment, the cursor was driven by a self designed glove based on an Ascension Technology Bird™. Only translations were involved in the fishing task. The graphics update rate was controlled at 15 Hz.

Experimental Design and Procedure

Eleven male and one female paid volunteers served as subjects in this experiment. The subjects were screened through the Bausch & Lomb Orthorator visual acuity and stereopsis tests. Subjects ages ranged from 18 to 36, with the majority in their early and mid-20’s. One of the 12 subjects was left handed and the rest were right handed, as determined by the Edinburgh inventory. Subjects were asked to wear the input glove on their dominant hand.

A balanced within subjects design was used. The 12 subjects were randomly assigned to a unique order of the four conditions (SS, SW, MS, MW) by a hyper-Graeco-Latin square pattern, which resulted in every condition being presented an equal number of times as first, second, third and final condition.

Following a 2 minute demonstration of all experimental conditions, the experiments with each subject were divided into four sessions, with one experimental condition in each session. There was a 1 minute rest between every two sessions. Each session comprised 5 tests. Test 1 started when the subject had no experience with the particular experimental condition. Test 2, 3, 4, and 5 started after the subjects had 3, 6, 9 and 12 minutes experience respectively. Practice trials occurred between the tests. Each test had 15 trials of fish catching. At the end of each test, the number fish both caught and missed (as both an absolute number and a relative percentage) and mean trial time were displayed to the subject.

At the end of the experiment, a short questionnaire was conducted to collect users' subjective preferences for all experimental conditions.

Performance Measures

Task performance was measured by trial completion time, error rate (capture and miss percentage) and error magnitude. Trial completion time was defined as the time duration from the beginning of the trial to the moment when the subject grasped. Error rate was defined as the percentage of fish missed in a test. Whenever a fish was missed, the error magnitude was defined as the Euclidean summation of x, y, z errors:

$$\text{Error Magnitude} = \sqrt{e_x^2 + e_y^2 + e_z^2}$$

Experimental Results

In total our experiment, which comprised 3600 experimental trials (i.e. 2 (cursor types) x 2 (display modes) x 12 (subjects) x 5 (tests) x 15 (trials per test)), with 3 performance measures per trial (i.e. trial completion time, error rate, and error magnitude), yielded 10,800 data points. Linear variance analysis was used to evaluate the statistical significance of the independent variables and their potential interactions for each of the three performance measures. Logarithmic non-linear transformations were applied to completion time and error magnitude scores in analyzing statistical significance, since residual analysis showed that these two measures were skewed towards the short fitted values. This section presents the primary results of the statistical analysis.

Trial Completion Time. The variance analysis indicated that cursor type (F(1, 3567) = 1148.5, p<.0001), display mode (F(1, 3567) = 630.3, p<.0001), cursor type and display mode interaction (F(1, 3567) = 253.5, p<.0001), subjects (F(11, 3567) = 96.1, p<.0001), learning phase (F(4, 3567) = 70.1, p<.0001) and trial number (different fish size) all very significantly affected trial completion time. Fig. 4 illustrates the effect of cursor type (silk vs. wire frame) and display mode (stereo vs. mono) to completion time.

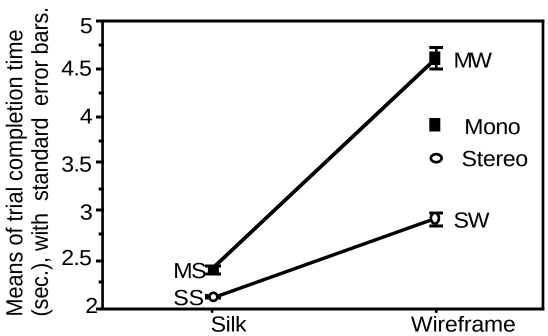

Figure 4: Trial completion time performance in relation to cursor type and display mode

Ranking these results in the order from best to worst, the mean completion time for each of the four interfaces were as follows. SS: 2.09 sec.; MS: 2.38 sec.; SW: 2.90 sec.; MW: 4.61 sec. Post hoc analysis shows that the differences between every pair of interfaces were significant, all at the p < .0001 level.

Error Rate. The statistically significant factors affecting error rate were: cursor type (F(1, 221) = 122.1, p<.0001), display mode (F(1, 221) = 67.9, p<.0001), cursor type and display mode interaction (F(1, 221) = 33.0, p<.0001), subjects (F(11, 221) = 5.75, p<.0001), and learning phase F(4, 221) = 3.69, p = 0.0062). Fig. 5 illustrates the effect of cursor type and display mode on error rate.

The performance order of the four interfaces measured in terms of error rate was exactly the same as that measured by trial completion time. The mean error rate for each combination of the four interfaces were SS: 12.7%, MS: 16.1%, SW: 20%, MW: 39.3%. The results of the Post-hoc pairwise comparison on error rate were as follows. SS vs. MS: p = .0795; SS vs. SW: p = .0002; SS vs. MW: p < .0001; MS vs. SW: p = .0479; MS vs. MW: p < .0001; SW vs. MW: p < .0001. The statistical significance of the difference between SS and MS was rather weak; however, all other pairwise differences were significant to greater degrees.

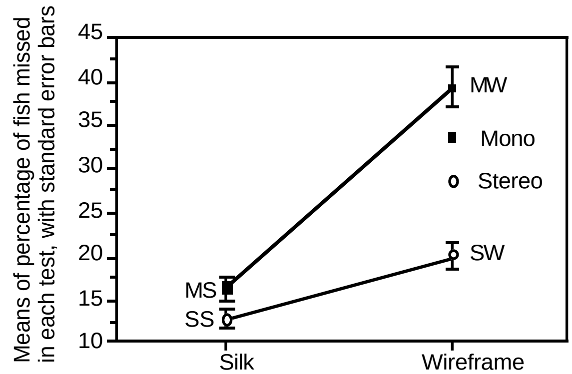

Figure 5: Error rate in relation to cursor type and display mode

Error Magnitude. The effect of cursor type and display mode on error magnitude are shown in Fig. 6. Error magnitude was significantly affected by cursor type (F(1, 761) = 19.9, p<.0001), display mode (F(1, 761) = 39.2, p<.0001), subjects (F(11, 761) = 3.60, p<.0001), and experimental phase (F(4, 761) = 3.88, p = 0.004). No significance for cursor type and display mode interaction (F(1, 761) = .009, p = .92) was found, however.

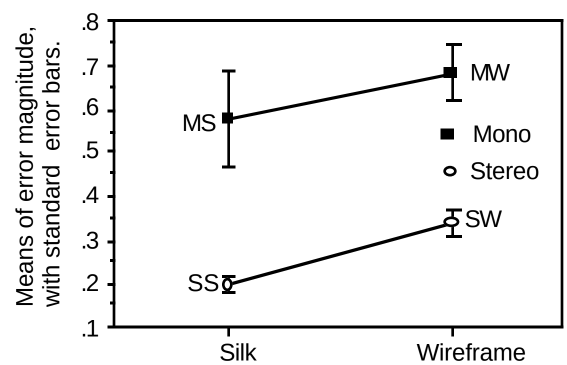

Figure 6: Error magnitude in relation to cursor type and display mode

The means of error magnitude were 0.197, 0.337, 0.577, and 0.68 for SS, SW, MS, and MW interfaces respectively. Note that, in contrast to the other two error measures, SW produced smaller errors than MS; however, the pairwise difference between SW and MS was not statistically significant (p = 0.16). All other pairwise differences were significant (from p = .003 to p < .0001).

Learning Effects. Fig. 7 demonstrates subjects' completion time performances in relation to the learning phase. It shows that the relative scores between the interfaces were consistent over the experimental tests. Subjects improved their time score in SS, MS and SW mode as they gained more experience, and presumably more confidence. Little improvement in completion time was found with the MW condition.

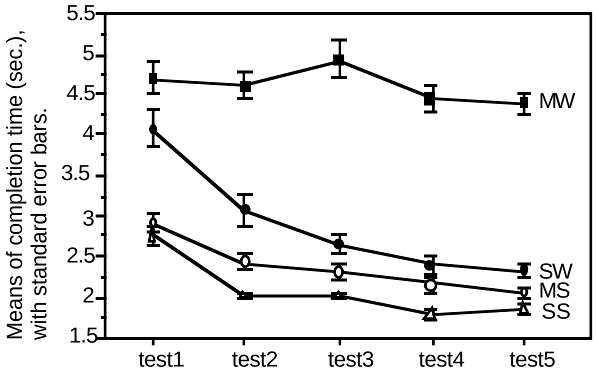

Figure 7: Time performance with four interfaces at each learning phase

Fig. 8 gives error rate in relation to learning phase. Again the relative rank of each mode was consistent across all five phases of the experiment. Interestingly, error rate for the MW condition showed the most obvious improvement over the experiment. A small amount of improvement was found in the MS condition, and essentially none in the SS and SW modes.

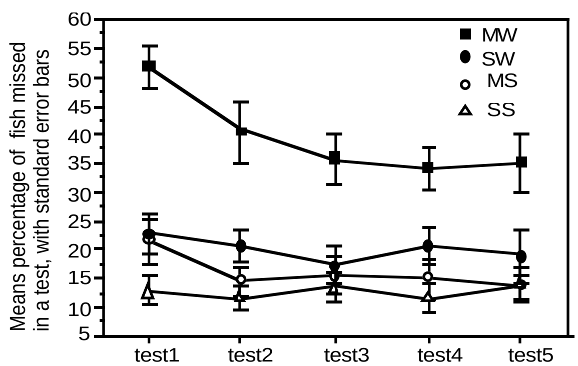

Figure 8: Error rate with four interfaces at each learning phase

Comparing Fig. 7 with Fig. 8 reveals important information about speed accuracy tradeoff patterns with respect to learning. For the MW mode, subjects had more than a 35% error rate, which apparently caused them to focus on improving the accuracy aspect of the task at the expense of time performance. In the other three cases (SS, MS, and SW), subjects already had less than a 25% error rate and it appears that they were more satisfied with this level of accuracy, and thus were devoting more effort to shorting their trial completion times.

Subjective Preference. Subjective evaluations (Table 1) are consistent with the other performance measures. SS was the most preferred and MW was the least liked. Of special interest is the fact that MS was ranked higher than SW.

Summary of Results. The experiment largely confirmed our initial hypothesis. In general, the "silk surface" was the most effective factor for successful acquisition. While stereo presentation in combination with the silk surface improved performance significantly, performance with the silk surface in mono display mode was in fact better than the stereo wire frame case for all measures except error magnitude, for which no significant difference was found.

Subjective Preferences

	Very Low	Low	OK	HIgh	Very High
MW	8	3	1		
MS			4	4	4
SW		2	7	3	
SS				3	9

Table 1: Subjective preferences (each cell contains the number of subjects with that rating)

DISCUSSION

Even though comparatively little perceptual research has been carried out on the relative strengths of various depth cues, of which only a small portion has addressed issues specifically related to computer graphic presentation, the results of our experiment appear to confirm some of those earlier investigations. In particular, in an early cue conflict study, Schriever [10] compared the relative influences of binocular disparity (i.e. stereoscopic displays), perspective, shading and occlusion, and showed among other things the dominance of occlusion over disparity information. More recently, Braunstein et al [1] showed that conflicting edge-occlusion dominates disparity. Even more recently, Wickens et al [14] in a review of depth combination literature, concluded that motion, disparity and occlusion are the most powerful depth cues in displays. The results presented here clearly contribute to that literature by illustrating some of the powerful advantages that can be afforded by augmenting visual feedback through either partial occlusion or binocular disparity, and in particular both in combination.

Two points of particular interest with respect to the silk cursor distinguish this research from other studies. One of these is the fact that the silk cursor does not block completely the view of any object which it occludes, due to the fact that it is *semi*-transparent. In essence, therefore, we contend that not only are important enhancements of depth perception to be gained through application of occlusion cues, but the one clear *dis*advantage of complete occlusion is greatly diminished – namely, the fact that all information about objects being obscured by an opaque intervening object is necessarily lost. For such practical computer-related applications as pursuit tracking, docking, target acquisition, etc., this is expected to present a significant advantage.

The second point relates to the fact that the silk cursor provides *discrete*, rather than *continuous*, levels of depth information. This is in contrast to stereoscopic displays, which are able to provide information not only about *whether* one object is farther away than another object, but also to a significant extent *by how much* they are separated in space, by means of binocular depth scaling. The silk cursor, on the other hand, is not able to provide this information. In Fig. 5 we see that the error rate for the MS case was lower than that of the SW case, which supports the effectiveness of the silk cursor as a discrete capturing device. However, upon examining Fig. 6 we note that the magnitude of errors for the MS case are larger than those of the SW case. The implication of this is that, although fewer errors were made with the silk surface cursor, the magnitude of those fewer errors must have been relatively larger than for the SW case, suggesting that continuous depth information was not being used.

In summary, we are quite encouraged by these results with the semi-transparent silk cursor, especially for applications in 3D interactive environments. In one existing application – our research on evaluating isometric versus elastic 6 DOF controllers – we were long hindered by the lack of an adequate display means which would allow us to concentrate on the control aspects of the experiment, even though we had already been using a stereoscopic display, the use of the silk cursor overcame the earlier display bottleneck, and allowed us to conduct that 6 DOF tracking experiment successfully [17].

FUTURE WORK

Of the two novel aspects of the silk volume cursor technique, this paper has focused on the "silk stocking" effect, i.e. the use of partial occlusion for enhancing depth cues. Our future work will include extending the classical Fitts' Law model to analyse the "Prince" technique and the effects of cursor versus target size.

CONCLUSION

We have proposed a semi-transparent silk volume cursor, to serve as a novel technique for performing target acquisition type tasks in 3D environments. Within the context of a carefully designed "virtual fishing" experiment that represented a dynamic 3D target acquisition task, the silk volume cursor demonstrated superior performance over a comparable wire frame cursor, both in stereo and in mono display modes.

ACKNOWLEDGMENTS

We would like to thank the members of the Input Research Group and the Ergonomics in Telerobotics and Control (ETC) Lab at the University of Toronto, who provided the forum within which this work was undertaken. Primary support for this work has come from the Information Technology Research Centre of Ontario, the Defence and Civil Institute of Environmental Medicine, the Natural Sciences and Engineering Research Council of Canada, and Xerox PARC. Additional support has been provided by Digital Equipment Corp. and Apple Computer Inc. This support is gratefully acknowledged. In addition, the authors would like to thank Ferdie Poblete for the initial drawing of the fish used in the experiment.

REFERENCES

1. Braunstein, M.L., Anderson, G. J., Rouse, M. W., Tittle, J. S., Recovering viewer-centered depth from disparity, occlusion and velocity gradients. *Perception & Psychophysics*. 40 (1986). 216-224.
2. Card, S., G. Robertson, and J. Mackinlay. The information visualizer. *Proc. of CHI* (1991). pp. 181-194.
3. Chen, M., S. J. Mountford, and A. Sellen. A study in interactive 3-D rotation using 2-D control devices. *Proc. of ACM Siggraph* (1988).
4. Evans, K., P. Tanner, and M. Wein. Tablet-Based Valuators That Provide One, Two, or Three Degrees of Freedom. *Computer Graphics*. 15(3) (1981). 91-97.
5. Fitts, P.M. The information capacity of the human motor system in controlling the amplitude of movement. *J. of Experimental Psychology*. 47 (1954). 381-391.
6. Foley, J. D., A. van Dam, S. K. Feiner, J. F. Hughes. *Computer Graphics Principles and Practice*. 1990, Reading, MA: Addison-Wesley.
7. MacKenzie, I. S., A. Sellen, and W. Buxton. A comparison of input devices in elemental pointing and dragging tasks. *Proc. of CHI* (1991). 161-166.
8. Mackinlay, J. D., S. Card, and G. G. Robertson. Rapid controlled movement through a virtual 3D workspace. *Computer Graphics, Proc. of SIGGRAPH*. 24(3) (1990). pp. 197-176.
9. Poulton, E.C., *Tracking skill and manual control*. 1974, New York: Academic Press.
10. Schriever, W. Experimentelle Studien über das stereoskopische Sehen. *Zeitschrift für Pscyhologie*. 96 (1925). 113-170.
11. SIGGRAPH. Proceedings of Workshop on Interactive 3D Graphics. (1986 - 1992). ACM SIGGRAPH.
12. Sollenberger, R. L. and P. Milgram. Effects of stereoscopic and rotational displays in a three-dimensional path-tracing task. *Human Factors*. 35(3) (1993). 483-499.
13. Venolia, D. Facile 3D direct manipulation. *Proc. of INTERCHI* (1993). pp. 31-36.
14. Wickens, C. D., S. Todd, and K. Seedier, *Three-dimensional displays: Perception, implementation and applications*. (1989). CSERIAC Technical Report 89-001. Wright Patterson Air Force Base, Ohio.
15. Yeh, Y. Y. and L. D. Silverstein. Spatial judgments with monoscopic and stereoscopic presentation of perspective displays. *Human Factors*. 34(5) (1992). 583-600.
16. Zhai, S. and P. Milgram. Human Performance Evaluation of Manipulation Schemes in Virtual Environments. *Proc. of VRAIS'93: IEEE Virtual Reality Annual International Symposium*. (1993). pp.155-161.
17. Zhai, S. and P. Milgram. Human performance evaluation of isometric and elastic rate controllers in a 6 DOF tracking task. *Proc. SPIE Vol. 2057 Telemanipulator Technology & Space Telerobotics* (1993).