The Importance of Stereo and Eye Coupled Perspective for Eye-Hand Coordination in Fish Tank VR

Roland Arsenault and Colin Ware

Data Visualization Research Lab
Center for Coastal and Ocean Mapping
University of New Hampshire

It is possible to simulate a high quality virtual environment with viewpoint controlled perspective, high quality stereo and a sense of touch obtained with the Phantom force feedback device using existing "fish tank VR" technologies. This enables us to investigate the importance of different depth cues and touch using higher quality visual display than is possible with more immersive technologies. Prior work on depth perception suggests that different depth cues are important depending on the task performed. A number of studies have shown that motion parallax is more important than stereopsis in perceiving 3D patterns, but other studies suggest that stereopsis should be critically important for visually guided reaching. A Fitts' Law tapping task was used to investigate the relative importance of stereo and head tracking in visually guided hand movements. It allowed us to examine the inter-tap intervals following a head movement in order to look for evidence of rapid adaptation to a misplaced head position. The results show that stereo is considerably more important than eye-coupled perspective for this task and that the benefits increase as task difficulty increases. Disabling stereo increased mean inter-tap intervals by 33% while disabling head tracking produced only a 11% time increase. However we failed to find the expected evidence for adaptation during the series of taps. We conclude by discussing the theoretical and practical implications of the results.

1 Introduction

When we reach for an object it is critical that we accurately judge the distance to that object. If we cannot make such judgments our interaction will become strained and take attention away from the primary task, which might be examining scientific data or constructing a virtual environment. In the present paper we address the relative importance of coupling perspective to the user's eye position and stereoscopic viewing, for rapid reaching in 3D virtual environments. To introduce the subject we review prior work on the importance of stereoscopic depth information, correct perspective and active touch in virtual reality and we also discuss the different display configurations that are appropriate for this research.

Research into depth perception is traditionally centered around depth cues such as stereoscopic depth, motion parallax, occlusion and perspective in providing distance information. The depth cues of stereoscopic disparity and motion parallax may be especially important for visualizing the positions of objects floating in space. Coupling the perspective to the user's eye position is often considered to be one of the defining characteristics of a VR system and this enables motion parallax information to be obtained when the user moves with respect to a virtual scene.

Stereoscopic depth is the information we gain from *disparities* – differences in relative separation between pairs of features imaged in the two eyes (Durgin, Proffitt, Olsen & Reinke, 1995; Patterson & Martin, 1992). Stereopsis has a variable utility as a function of distance from the observer. Disparities become too small to be useful for objects at great distances because the images in the eyes become essentially identical. Disparities can also be too large, resulting in *diplopia* (double images), and this occurs when objects differ too much in depth. Diplopia can occur with disparities as small as 10 minutes of arc in the worst case (Howard & Rogers, 1995). For images viewed in stereo on a computer screen, roughly at arms length, this translates into only a few centimeters of usable relative depth. However, a number of factors such as relative motion, and depth of focus can enable us to fuse images with greater depth (Patterson and Martin, 1992). In general, stereo is a very strong cue for judging the relative depth of nearby objects that are close to being equidistant from us, but it is a poor cue for judging large depth differences.

Another aspect of stereopsis is that it is a *super-acuity*, meaning that we can resolve very small differences, smaller indeed than can be predicted on the basis of the spacing of receptors in the eye (Howard & Rogers, 1995). We can resolve disparities as small as 10 seconds of arc. However most head-mounted displays have very large pixels (e.g. 800 pixels horizontally spread out over more than 90 degrees of visual angle) and are therefore only capable of generating disparities greater than 6-8 minutes of arc. This means that only two or three depth steps are displayable in certain circumstances before double imaging occurs. Admittedly this is the worst case, and anti-aliasing and other factors can improve the situation, but nevertheless stereopsis is likely to be more useful with either much higher resolution screens or screens that concentrate the pixels into a smaller visual field. This suggests that, given current monitor technology, small field-of-view displays are best to study stereoscopic depth-related phenomena.

Motion parallax refers to the depth information that is obtained as an observer moves relative to the environment. A number of studies have compared the value of motion parallax and stereopsis and the results appear to depend on the task. Ware, Arthur & Booth (1993) and Hendrix & Barfield (1996) both reported that motion parallax from head coupled perspective increased the sense of realism or "presence" of the virtual environment. However, Ware et al found that motion parallax was the more important factor whereas Hendrix and Barfield found them to be about the same. For a surface orientation perception task Norman, Todd, and Phillips (1995) found that both stereopsis and motion parallax helped in the perception of the surface orientation to roughly the same extent. Bradshaw, Parton, and Glennerster (2000) found that the relative value of stereopsis and parallax reversed from near viewing at 150 cm to far viewing at 300 cm for a triangle-matching task. Motion parallax was the more important cue in the near viewing condition but was not as useful in the far condition.

For the task of tracing paths in network or tree structures, a number of studies have shown that motion parallax is a more important cue than stereoscopic depth (Sollenberger & Milgram, 1993; Ware & Franck, 1996). For example, for the task of path finding between nodes Ware and Franck showed that networks approximately 120% larger could be viewed with motion parallax information compared to a static view. Stereoscopic depth only provided a 60% increase in the size of the network that could be viewed.

Stereoscopic depth may be the more valuable cue for visually guided reaching. Stereoscopic depth has been shown to dramatically improve performance for 3 dof pick-and-place tasks (Kim, Tendick & Stark, 1993), but this study did not include head coupled perspective. Lion (1993) investigated both stereoscopic depth and head coupled perspective for a task in which the subject had to move a ring along a wire curved in 3D space. He found a large advantage from stereoscopic viewing but none from head tracking. In a similar study Boritz and Booth, (1997) used a 3D point location task and found that stereo viewing increased performance substantially both in accuracy and task completion time, whereas head-coupled perspective again had no effect. However, in both studies there was nothing in the task requiring that subjects change their viewing position. As a consequence, subjects may have carried out the experiment from more-or-less a single viewpoint – in which case it would be hardly surprising that coupling perspective to head position had little effect. The study we report here is similar to theirs with the important difference that in our task subjects had to make substantial changes in viewing position.

There are a number of reasons why we might expect stereopsis to be more important than motion parallax for near-field reaching tasks. One is the simple observation that people who do fine positioning tasks, such as threading a needle, hold their heads steady and therefore do not appear to use parallax information (Stoffregen, Smart, Bardy, & Pagulayan, 1999). In addition, Bingham, Bradley, Bailey, and Vinner (2001) have suggested that disparity matching is critical for calibrating eye-hand coordination for visually guided reaching. As they put it, "It is often assumed that the guidance of reaching is the ultimate function of binocular vision".

1.1 Perspective Distortion

Even if subjects do not move their heads much when carrying out reaching tasks (and therefore do not generate motion parallax), there is still a good reason for tracking head position. For every perspective image there is a point called the center of perspective from which the image should be viewed for the perspective to be correct. When an image is viewed from a different point, geometry suggests that distortions should be perceived (see Figure 1). Moreover, if the task is visually guided hand movement in VR, then the image of whatever represents the hand will become displaced from the actual position of the hand. To make this worse, the amount of displacement will be a function of the distance behind the virtual screen.

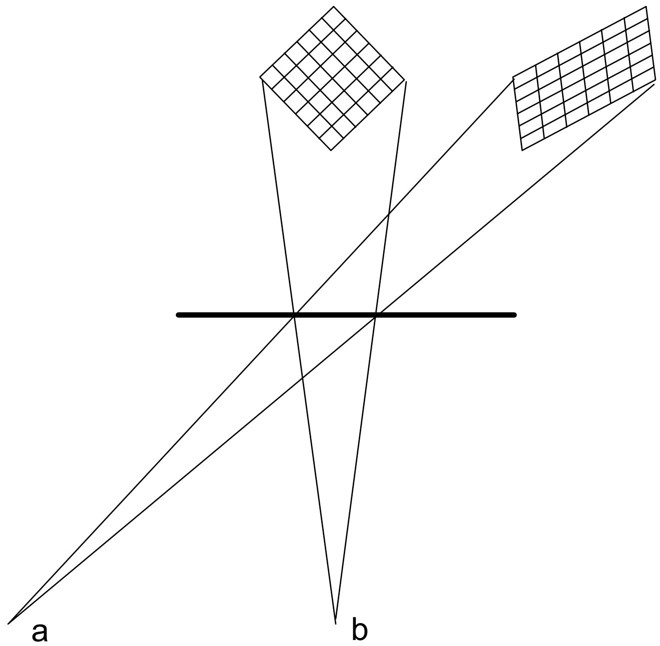

Figure 1: If an image is computed with a center of perspective at b and viewed from location a, then geometry dictates considerable distortion, assuming depths are properly perceived.

On the other hand there is also evidence that we are mostly insensitive to the large distortions that geometry predicts. Few people are aware of perspective distortion when they are watching movies or television, even though they may be doing so from a radically incorrect viewpoint relative to the center of perspective. This lack of sensitivity is sometimes called the robustness of linear perspective (Kubovy, 1986). One of the mechanisms that can partially account for the lack of perceived distortion may be based on a built-in perceptual assumption that the world is rigid. Studies with subjects wearing prisms also tell us that we can rapidly adapt to prism displacement of the seen hand position relative to the felt hand position and this may mean that distortions of the relative position due to failure to track head position may not be important in reaching tasks (Rosetti, Koga, & Mano, 1993). However, there is a considerable difference between the kind of distortion that occurs with a prism and the geometric changes resulting from an incorrect viewing position. Hence, a primary goal of the present experiment was to allow us to look the effects of incorrect perspective with stereoscopic viewing for visually guided reaching.

1.2 Active Touch

Although vision may be the primary channel with which we take information about the world, the sense of touch can tell us about properties of objects, such as surface roughness and elasticity (Klatsky & Lederman, 1999) and we can also feel constraints that are useful for object positioning. Thus, for example, when we place an object on a table top, gravity constrains the task to three degrees of freedom (two for position, one for orientation) making it much easier than an unconstrained six degree of freedom positioning task (Wang & MacKenzie, 2000). Touch may also be important in 3D visually guided reaching because touching an object may cause a re-calibration of the stereoscopic depth estimation system (Bingham, Bradley, Bailey & Vinner, 2001). Many previous studies of 3D reaching have been carried out in environments where a sense of touch is not simulated (e.g. Ware & Balakrishnan,1994; Boritz & Booth, 1997) and this may have lead to results that do not directly apply to environments where touch information is available. Force has been shown to help in peg-in-hole tasks (Sheridan, 1992) and in a tapping task, similar to the one reported here. We (Arsenault & Ware, 1999) found that enabling subject to feel the surfaces they were tapping increased the rate at which they could perform the task by about 12%.

1.3 Fish Tank VR as a Research Platform

The term “fish tank VR” describes a method for creating a small high-quality virtual environment (Schmandt, 1983; Deering, 1992; Ware, Arthur & Booth, 1993). By having a small field of view with a high-resolution monitor it is possible to get reasonable quality stereoscopic depth. By tracking the user’s head position it is possible to get the motion parallax that results from natural head movements with respect to a static object. In addition, errors in head-orientation tracking result in much smaller relative positioning errors for virtual objects compared to the case with a head-mounted display. A number of configurations have been studied. In the earliest, Schmandt (1983) used a semi-transparent mirror so that the user could see their hand with stereoscopically viewed 3D graphics imagery. One of the problems with the semi-transparent mirror used is that the occlusion depth cue is not preserved; the hand is seen transparently through solid objects and this can interfere with depth perception. Other versions (Ware, Arthur & Booth, 1993; Deering1992) used head-tracked stereoscopic glasses and a directly viewed monitor to create the 3D virtual image. Still more recent versions, illustrated in Figures 2 and 3, have used an opaque mirror that enables users to place hands in the virtual workspace and this provides an excellent platform for studies of eye-hand coordination (Wang et al, 1998; Ware & Rose, 1999). The mirror configuration makes it possible to manipulate the relationship between what is seen and what is touched. Using this kind of setup, Ware and Rose found that having the hand and the object co-located speeded object rotations, compared to the situation when the input device was held in a different spatial location.

2 Evaluation of Correct Perspective and Stereoscopic Depth for Tapping Task

Our goal in the present study was to examine the effects of correct versus incorrect perspective and stereoscopic depth for visually guided reaching. We were also interested in the time course of any adaptation that might occur when subjects changed their viewpoint, especially under conditions of perspective distortion when the head position was not tracked.

We chose a task that could be performed rapidly - a variation on the classic Fitts tapping task (Fitts, 1954). In a Fitts’-Law experiment, the time to reach a target is measured with distance to the target and target width as independent variables. The methodology has been used in hundreds of studies both as a pragmatic tool for comparing devices and as a means for evaluating theories of visually guided reaching. Using the Fitts'-Law method provides a link to this substantial body of empirical data and theory.

In Fitts' original task, subjects tapped back and forth between two strips of metal and he varied the width of the two strips and the distance between them. In many experiments it has been shown that the resulting data can usually be closely approximated by a simple function (Fitts' Law):

$$MT = C_1 + C_2 \log_2(D/W + 1),$$

where MT is the mean movement time, D is the distance to the target and W is the width of the target. The expression (D/W + 1) is called the index of difficulty. Note that there are many variations on the way the index of difficulty is calculated. We chose this one for reasons that are explained in Mackenzie (1992). C_1 and C_2 are empirical constants typically obtained from studies involving hundreds of trials and many subjects. A useful derivation from a Fitts'-Law calculation is $1/C_2$, which is called the Index of Performance.

In our variation on the Fitts' experiment, we had subjects tapping from the top of one cylinder to another. We varied the diameters of the cylinder tops and the distances between them to give a number of index of difficulty values (see Figure 4). We designed a task where the subjects tapped a whole set of targets in series to give a sequence of inter-tap intervals. By looking at the time course of the inter-tap intervals, we hoped to be able to learn about how rapidly subjects could adapt to a change in head position. When head position tracking was not used to set perspective, we expected to find evidence of improvement over the sequence of taps as subjects adapted to the incorrect perspective.

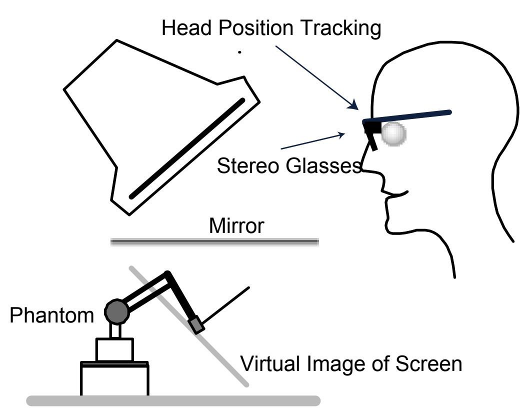

Figure 2: Schematic diagram of the apparatus used in the study. See text for explanation

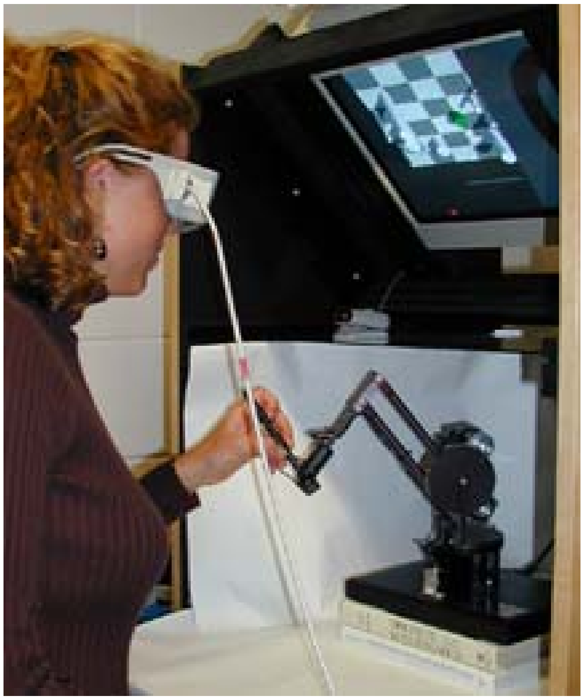

Figure 3. The various components of the apparatus are shown. The subjects looked down into the mirror to see the virtual image. The phantom has been moved forward for clarity.

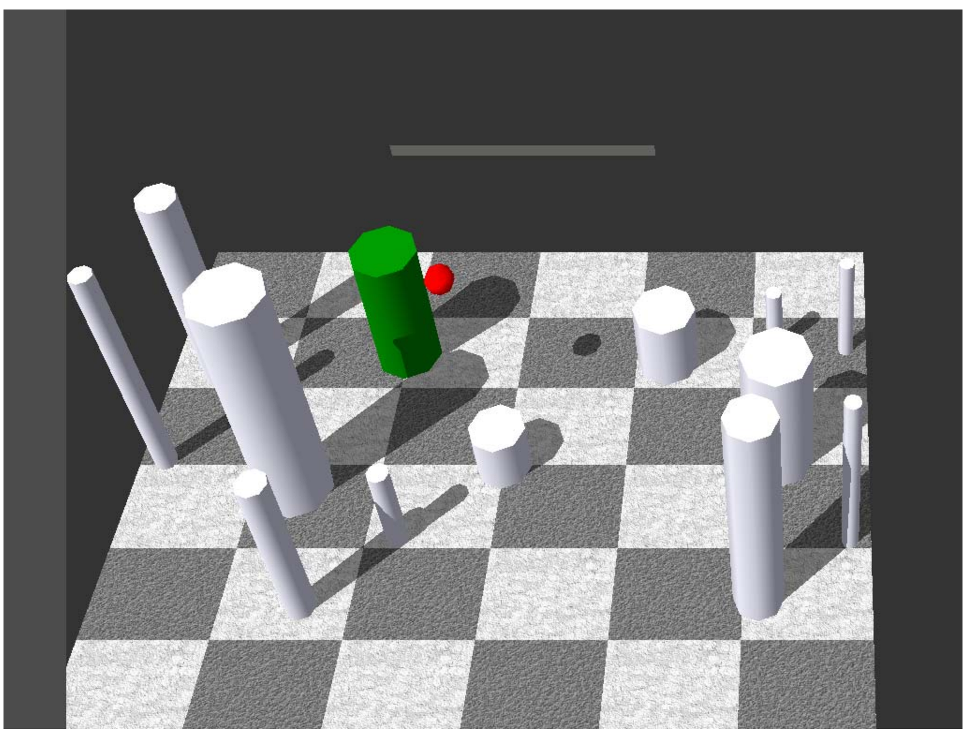

Figure 4. A set of virtual targets. A new set of virtual targets was generated for each trial block. The dark grey patch on the left is the edge of the virtual barrier that required subjects to shift head position.

2.1 Apparatus and Virtual Workspace

A virtual environment with coincident haptic and visual display was used for this study. The apparatus is illustrated in Figures 2,3 and 4. It contained a mirror mounted horizontally with a monitor mounted above it at a 45 deg angle as shown. This enabled the subject to place his or her hand in the virtual workspace.

A Phantom 1.0 from Sensable Technologies was used to provide a haptic workspace measuring12.7 cm x 17.8 cm x 25.4 cm. The Phantom has a mechanical jointed arm that both tracks the position of a hand held stylus and can provide an arbitrary force to the tip of that stylus (Massie & Salisbury, 1994). We rotated and translated the visual coordinate system so that it became coincident with the Phantom coordinate system. A 3D cursor consisting of a red sphere showed where the tip of the Phantom Stylus was located. Cast shadows were rendered for all objects and provided an additional depth cue for the cursor.

LCD shutter glasses were used to provide a frame sequential stereoscopic display. In all conditions the monitor refresh rate was 120 frames per second. Head tracking was achieved by attaching a sensor from a Polhemus 3Space Fastrack to the side of the stereo shutter glasses. By tracking the position and orientation of the shutter glasses the position of each eye was estimated and this information was used to provide a correct perspective image to each eye (Deering, 1992).

Calibration of the virtual workspace was verified by replacing the mirror with clear glass. This allowed faint computer graphics imagery to be superimposed on a physical object having the same dimensions and location. When properly calibrated, the virtual and physical objects remained co-registered for an observer despite changes in viewpoint.

2.2 Task

The subjects' task was to tap the tops of a series of cylinders of differing sizes. These cylinders were arranged on top of a checkerboard ground plane as illustrated in Figure 4. As soon as a cylinder was tapped its color was changed to white and the next in the series was highlighted red. A virtual barrier was introduced into the workspace above the targets and to one side. This forced the subject to lean to the right or the left to look around the barrier. The objective was to force a change in the subject's head position. The side of the barrier was alternated for each trial block. The virtual barrier extended from the midpoint (above the target space) to the right or to the left of center as illustrated in Figure 5.

2.3 Conditions

There were three independent variables as follows:

Stereo vs no Stereo [S; noS]

In the stereoscopic condition, alternate frames provided different images to the two eyes with the aid of shutter glasses. In the no-stereo condition, subjects saw the same image with both eyes. For the head-tracked condition, the viewpoint for both eyes was based on the midpoint between the two eyes.

Head tracked vs fixed perspective [HT ; noHT]

In the head-tracked condition, the center of perspective was based on the calculated eye position for each eye. In the non head-tracked condition a default center of perspective was based on the (roughly) estimated midpoint of the normal range of head movement (with lateral offsets for stereoscopic viewing). Note that even in the fixed perspective condition, the virtual barrier still took head position into account (even though the perspective view of the targets did not), forcing subjects to move their heads to one side or another.

Index of difficulty (ID)

Four values of index of difficulty were used. The distance between targets and the sizes of the targets were varied to produce the values 2, 3, 4 and 5.

2.4 Trials

Trials were carried out in sequences of 12 as the subjects tapped from target to target with the end to one trial triggering the start of the next. We were interested in the time course of the inter tap interval over the course of a trial block.

There were 13 target cylinders consisting of a home target and 12 others generated as described below. A single trial consisted of tapping one of the targets. A trial sequence consisted of tapping all 13 targets beginning with the home target. Following the tap on the home target, all trials were timed. The sizes and positions of the 12 trial targets were constructed to produce three replications of each of the 4 index of difficulty values.

We carried out 1 practice trial sequence in each of the four conditions noHT/noS, noHT/S, HT/noS, and HT/S, followed by 5 further replications of the set in a different random order. Because each trial sequence yields 3 instances of each index of difficulty, the 5 replications of the experiment yielded 15 values per subject for each of 16 conditions [HT (2) x Stereo(2) x Index-of-difficulty (4)].

Between each trial sequence subjects were required to move their viewpoint either to the left or the right. This was enforced by a movement of the barrier as illustrated in Figure 5. In the case of the noHT condition the barrier still moved but the perspective view of the test environment did not change. The amount of movement required to see around the barrier was about +/- 8 cm and this corresponded to an off axis viewing angle between 10 to 15 degrees to one side or the other with respect to the center of the test environment.

Once they had made this head movement subjects initiated a new trial sequence by depressing the space bar.

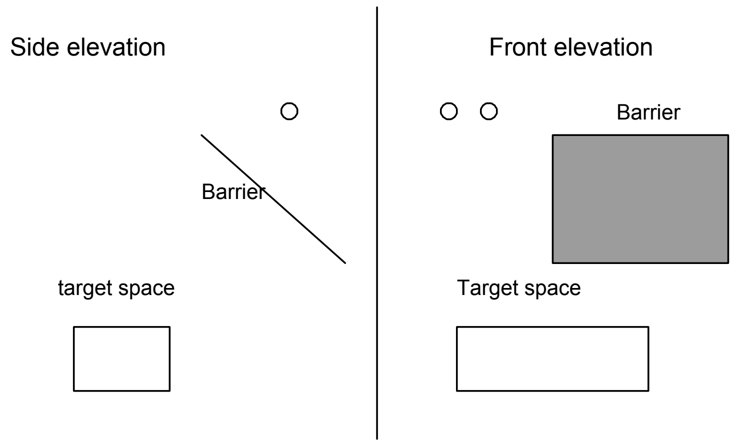

Figure 5. A barrier was used to force the subjects to view the target space from one side or the other.

2.5 Target Generation Algorithm

Targets were all generated within a bounding box extending 25 cm wide, 8 cm high and 11 cm deep. The Phantom neutral position is 1.0 cm below the center of this box. Physical constraints of the Phantom made it hard or impossible to reach the corners of the described workspace. Positions were therefore also constrained by an ellipsoid approximating the reach of the Phantom. We call the intersection of the box and the ellipsoid the *target space*.

For a particular trial block (of 12 trials) the first step involved randomizing the sequence in which the index of difficulty values would be given (3 each of IDs 2,3,4,5)
The first target was placed 4 cm below in the center of the workspace (this made the top of the cylinder coincided with the floor of the bounding box).

The following is the algorithm used to create sequences of 12 targets:

1. Select an index of difficulty, from the pre-computed random sequence.
2. Randomly find a position within the target space.
3. Calculate the diameter of the top of the target using the distance from the previous target and the index of difficulty provided.
4. Reject target if it lies outside of the range of diameters of 0.5 to 2.0 cm. Also reject target if it is closer than 1cm from any previously defined target. In these cases, repeat steps 1,2 and 3 until an acceptable position/size combination is found.
5. Repeat steps 1,2,3 and 4 until all 12 targets are generated.

It is possible with this algorithm to produce an incomplete set of targets, with no more suitable positions to place the remaining targets. This situation is detected by counting how many attempts are made to place a new target. If the count reaches 10,000 the set of targets is rejected and we start over.

2.6 Subjects

There were 19 subjects, each of whom was a paid volunteer. Four of the subjects were lab members while the remainder were undergraduate students.

3 Results

We carried out an analysis of variance including head-tracking, stereo, index of difficulty and trial number. All of these factors were significant.

The main effect of head tracking was to reduce the average time per tap from 2.02 sec to 1.85 sec. $F(1,18) = 17.8$, $p < 0.0001$. This makes the average time about 11% longer without head tracking.

The main effect of stereo viewing was to reduce the average time per tap from 2.18 sec to 1.69 sec. $F(1,18) = 125.5$, $p < 0.0001$. Thus the average time was about 33% longer with stereo disabled.

There was a main effect of index of difficult, $F(3,54) = 1000$, $p < 0.000001$, and there was a significant interaction between index of difficulty and stereo, $F(3,54) = 17.1$, $p < 0.001$. This is illustrated in Figure 6, which shows that the advantage of stereo increases as the task difficulty increases.

Another interaction was found between index of difficult and head tracking, $F(3,54) = 7.5$, $p < 0.001$. This is illustrated in Figure 7 which shows that the advantage of head tracking also increases as task difficulty increases.

There was a main effect of trial number $F(11,18) = 7.5$, $p<0.001$. The average inter tap interval decreased over the course of the trials.

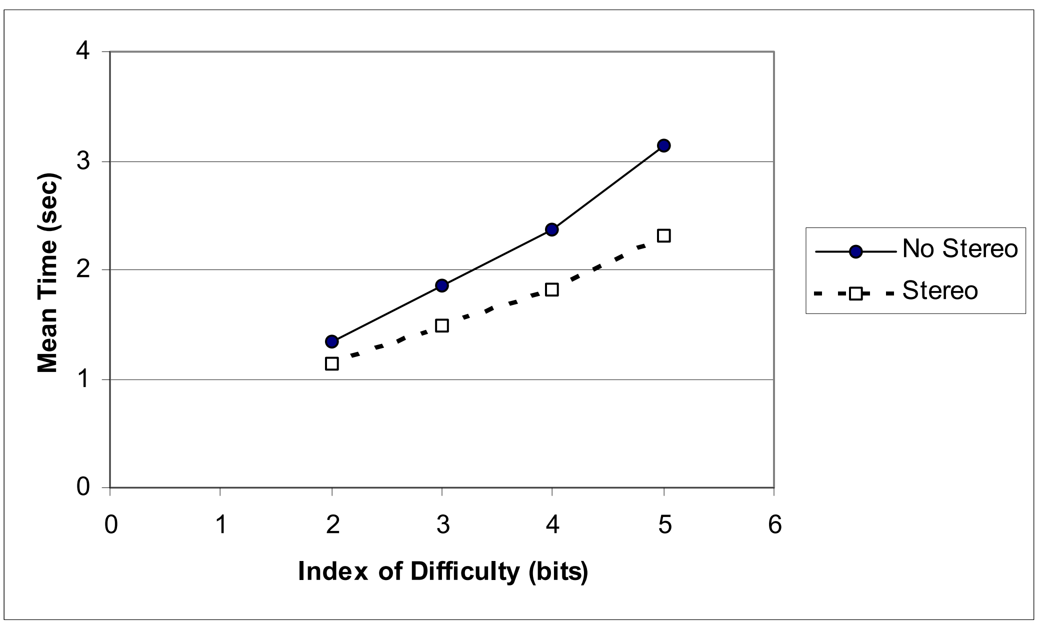

Figure 6. The mean time per tap is plotted against index of difficulty with and without stereo viewing.

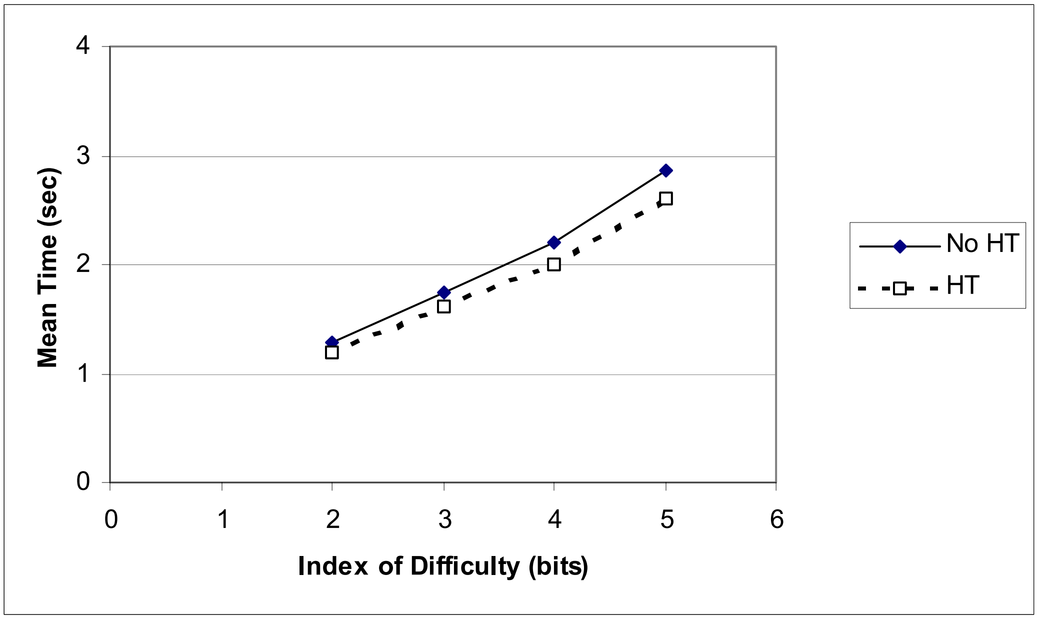

Figure 7. The mean time per tap is plotted against index of difficulty with and without head-tracked perspective.

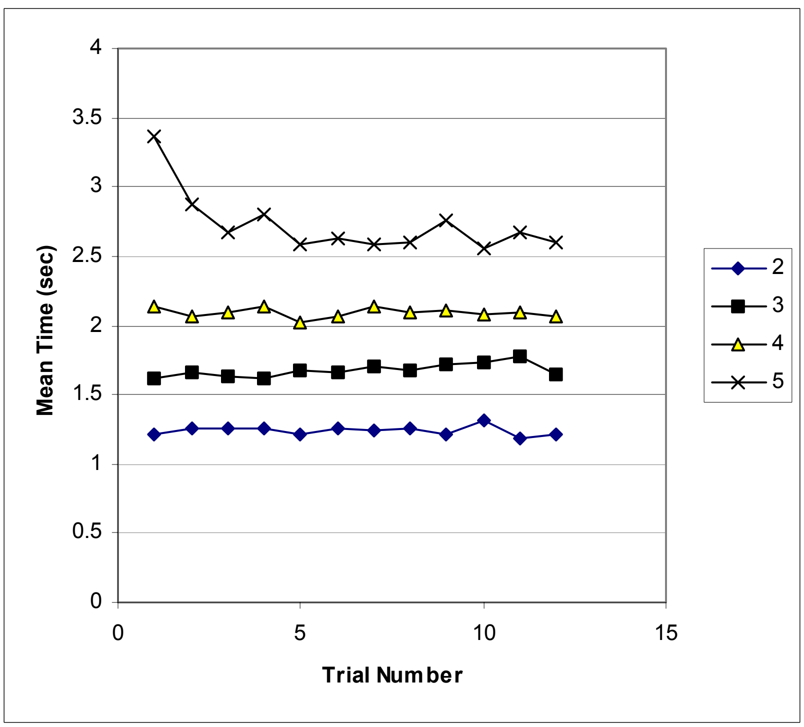

Figure 8. The time course of the inter-tap intervals is plotted as a function of trial sequence number for the different index of difficulty values.

There was a significant interaction between the trial number of a tap in the sequence of 12 inter-tap intervals and the index of difficulty [$F(33,594)=6.4$]. These data are illustrated in Figure 8. This shows an improvement over the first few trials for only the most difficult (ID=5) condition. However, there was no evidence to support our prediction that head tracking would enable a more rapid adjustment after a change of head position. Whether or not perspective was coupled to the eye position had no effect on the time course of the inter-tap intervals.

Our experiment allowed us to measure the Fitts Law index of performance benefits for stereo and head tracking. The IP values for the 4 conditions are given in Table 1. Overall, the gain in performance from including both head tracking and stereo was from 1.58 to 2.70 bits per second; more than 50% of this was attributable to stereo whereas 11% was attributable to eye-coupled perspective.

	No HT	HT
No Stereo	1.58	1.75
Stereo	2.41	2.70

Table 1: The index of performance values are shown for the four main conditions. Units are bits per second.

7 Discussion

Our results suggest that stereoscopic viewing is more important than eye coupled perspective for visually guided reaching tasks, with benefits that increase as the targets get smaller. The gain from linking perspective to eye position was relatively small but ideally head tracking should also be used since it also measurably improved performance.

Overall, our results add support to the growing evidence that the value of different depth cues differs from task to task (Bradshaw, Parton & Glennerster 2000). Our finding of the greater importance of stereoscopic depth contrasts with prior results from tasks such as tracing cerebral arteries and veins (or other 3D networks) that showed motion parallax obtained from head movement to be the more important depth cue (Sollenberger & Milgram, 1993; Ware & Franck, 1996).

Both in the present study and our previous investigation (Arsenault & Ware, 2000) we found that head tracking had a measurable effect on performance, whereas others (Lion, 1993; Boritz & Booth, 1997) found no effect. The most likely reason for this is that we created a task for which head movements were required whereas they did not. In some virtual reality tasks, looking around obstacles would be a normal part of interaction. In others it is likely to occur infrequently. Hence the value of this observation would also depend on the task mix.

We were surprised by the lack of a clear improvement over the first few taps that we could attribute to head position tracking. Rosetti, Koga, & Mano (1993) found substantial accuracy improvements over the first ten trials in a pointing task after prism displacements. We expected much the same. The reason why they found adaptation and we did not may lie in a more detailed examination of the task. Rosetti et al's experiment required ballistic hand movements. Subjects could not adjust hand position during the course of a trial because they were not able to see their finger at the start of the trial and after the trial strarted they made rapid (<200 msec) movements, too short for significant feedback to have occurred. In our study, adaptation to the misplaced position of the virtual probe could have been taken place when the subject moved to place the probe on the start object since continuous visual feedback of hand position was available. Also, Bingham, Bradley, Bailey & Vinner (2002) suggested that the contact of the hand with a target can cause recalibration of stereoscopic disparity information. Thus, recalibration may have occurred during the (unmeasured) interval in which the subject moved his or her hand into contact with the first target.

One practical consequence of our findings is that for fish tank VR accurate co-registration of eye and hand coordinate spaces may be unnecessary at least when the most common task is reaching for targets. Even the quite large discrepancies that occurred when head position was not tracked resulted in only small performance decrements. However, this should not be taken as evidence that accurate head tracking is not needed for other VR setups. In immersion VR, simulator sickness is likely to increase if accurate viewpoint estimation is not used.

Even though fish tank VR is quite unlike the wide-field experience obtained with a CAVE or an HMD the results may generalize to immersion VR, particularly for reaching tasks when the body is held static for fine positioning. Using fish tank VR as a research platform may be especially useful in studies of the value of stereoscopic display. The relatively small pixels allow for better stereoscopic depth information and hence can provide a better understanding of the potential value of this depth cue when high quality stereo becomes available for immersion VR systems.

Acknowledgements.
This work was funded through NSF grant IIS 0081292 to Colin Ware.

References

Arsenault, R.., & Ware, C. (2000). Eye-Hand Co-ordination with Force Feedback. *ACM CHI'2000 Proceedings,* 408-414.

Bimber, O., Encarnacao, L.M., & Branco, P. (2001). The Extended Virtual Table: An Optical Extension for Table-Like Projection Systems. *Presence, 10(6),* 613-631.

Bingham, G.P., Bradley, A., Bailey, M., & Vinner, R. (2001). Accommodation, Occlusion and Disparity Matching are used to guide reaching: A comparison of actual versus virtual environments. *Journal of Experimental Psychology: Human Perception and Performance, 27 (6),* 1314-1334.

Boritz, J., & Booth, K.S. (1997). A study of Interactive 3D point location in a computer simulated virtual environment. *ACM VRST'97,* 181-187.

Bradshaw, M.F., Parton, A.D., & Glennerster, A. (2000). The task-dependent use of binocular disparity and motion parallax information. *Vision Research, 40,* 3725-3734.

Cruz-Neira, C., Sandin, D., & DeFanti, T. (1993). Surround-screen projection-based virtual reality: The design and implementation of the CAVE. *Computer Graphics, Proceedings of SIGGRAPH'93,* 135-142.

Deering, M. (1992). High Resolution Virtual Reality. *Proceedings of ACM SIGGRAPH'92 Computer Graphics, 26(2),* 195-202.

Durgin, F.H., Proffitt, D.R., Olsen, J.TY. & Reinke, K.S. (1995). Comparing depth from motion with depth from binocular disparity. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 679-699.

Fine, I., & Jacobs, R.A. Modeling the combination of motion, stereo and vergence angle cues to visual depth. *Neural Computation, 11,* 1297-1330.

Fitts, P.M., (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology, 47(6),* 381-391.

Graham, E.D., & Mackenzie, C.L. Physical versus virtual pointing. *Proceedings of ACM CHI'96,* 292-299.

Hendrix, C. & Barfield, W. (1996). Presence within virtual environments as a function of visual display parameters. *Presence: Teleoperators and Virtual Environments, 5,* 275-289.

Howard, I.P., & Rogers, B.J. (1995). Binocular Vision and Stereopsis. *Oxford psychology series, no. 29.* Oxford University Press: Oxford.

Kim, W.S., Tendick, F., & Stark, L. (1993). Visual enhancements in pick-and-place tasks: Human operators controlling a simulated cylindrical manipulator. In S. Ellis (Ed.), *Pictorial Communication in virtual and real environments,* (pp. 182-195). London: Taylor and Francis.

Lion, D.M. (1993). Three Dimensional Manual Tracking using a head-tracked stereo display. (Technical Report TH-93-7). Human Interface Technology Lab.

Mackenzie, I.S. (1992). Fitts' Law as a research and design tool in human-computer interaction. *Human-Computer Interaction, 7,* 91-139.

Massie, T.H., & Salisbury, J.K. (1994). The PHANTom haptic interface: A device for probing virtual objects. *Proceedings of the ASME Dnamic Systems and Control Division, 55(1),* 295-301.

Norman, J.F., Todd, J.T. & Phillips, F. (1995). The perception of surface orientation from multiple sources of optical information. *Perception and Psychophysics, 57(5),* 629-636.

Patterson, R., & Martin, W.L. (1992). Human Stereopsis. *Human Factors, 34(6),* 669-692.

Rosetti, Y., Koga, K., & Mano, T., (1993). Prismatic displacement of vision induces transient changes in the timing of eye-hand coordination. *Perception and Psychophysics. 54(3),* 355-364.

Schmandt, C. (1983). Spatial input/display correspondence in a stereoscopic computer graphics workstation. *Computer Graphics, 17(3),* 253-261.

Sheridan, T.B. (1992). *Telerobotics, Automation and Human Supervisory Control,* MIT Press: Cambridge Mass.

Sollenberger, R.M. & Milgram, P. (1993). Effect of Stereoscopic and Rotational display in three-dimentional Path-Tracing task. *Human factors, 35(3),* 483-499.

Stoffregen, T. A., Smart, L. J., Bardy, B. G., & Pagulayan, R. J. (1999). Postural stabilization of looking. *Journal of Experimental Psychology: Human Perception & Performance, 25,* 1641-1658.

Wang, Y., MacKenzie, C.L, Summers, V.A., & Booth, K.S. (1998). The Structure of Object Transportation and Orientation in Human-Computer Interaction. *Proceedings of ACM CHI'98,* 312-319.

Wang, Y., MacKenzie, C.L. (2000). The role of contextual haptic and visual constraints on object manipulation in virtual environments. *Proceedings of ACM CHI'2000,* 532-538.

Ware, C., Arthur, K., & Booth K.S. (1993). Fish tank virtual Reality, *Proceedings of ACM INTERCHI,* 37-42.

Ware, C., & Balakrishnan, R. (1994). Object Acquisition in VR displays: Lag and Frame Rate. *ACM Transactions on Computer Human Interaction, 1(4),* 331-357.

Ware, C., & Franck, G. (1996). Evaluating Stereo and Motion Cues for Visualizing Information Nets in Three Dimensions. *ACM Transactions on Graphics, 15(2),* 121-140.

Ware, C., & Rose, J. (1999). Rotating virtual objects with real handles. *ACM Transactions on CHI, 6(2)*, 162-180.