# Lateral Head Tracking in Desktop Virtual Reality

B.R. Boschker and J.D. Mulder

Center for Mathematics and Computer Science CWI, Amsterdam, The Netherlands

### Abstract

Head coupled perspective is often considered to be an essential aspect of stereoscopic desktop virtual reality (VR) systems. Such systems use a tracking device to determine the user's head pose in up to six degrees of freedom (DOF). Users of desktop VR systems perform their task while sitting down and therefore the extent of head movements is limited. This paper investigates the validity of using a head tracking system for desktop VR that only tracks lateral head movement. Users performed a depth estimation task under full (six DOF) head tracking, lateral head tracking, and disabled head tracking. Furthermore, we considered stereoscopic and monoscopic viewing. Our results show that user performance was not significantly affected when incorporating only lateral head motion. Both lateral and full head tracking performed better than the disabled head tracking case.

Categories and Subject Descriptors (according to ACM CCS): H.5.0 [Information Interfaces and Presentation]: General I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism

#### 1. Introduction

Head coupled perspective is often considered to be an essential aspect of virtual reality (VR) systems. Head coupling gives the user a geometrically correct view of the scene, as well as enabling head movement-induced motion parallax as a depth cue. Several researchers [Dee92, McK92, WAB93, WF96] have shown motion parallax to be of great importance, improving both user performance in three-dimensional tasks and giving the user a better sense of depth than with stereopsis alone.

Systems for desktop virtual reality incorporating trackers that deliver head pose in six degrees of freedom can provide the user with geometrically correct images. When a system employs a tracking device that delivers head pose in less than six degrees of freedom, the images delivered to the user's eyes will most likely not be geometrically correct and therefore the visual space will appear distorted [WHR99, Pfa96, Pfa00].

We divide VR systems in two groups, based on the type of display system: systems where the display moves with the user's head (Head Mounted Displays, HMDs) and stationary display systems, where user movement is relative to the display. In the former case, small variations in head movements result in large changes in displayed virtual content. Therefore, the need for tracking head pose in six degrees of

© The Eurographics Association 2004.

freedom (DOF) is clear. The latter group can be subdivided in two classes, the first class being immersive virtual reality (e.g. CAVE [CNSD93]). This class is characterized by large projection-based displays and relatively large freedom of movement for the user. The second class of stationary display systems we distinguish, is referred to as 'Fish Tank' [WAB93] or desktop VR. This paper focuses on these types of systems.

Compared to other VR systems, the benefits of desktop VR systems are high display resolution and relatively low cost. Desktop VR systems are usually based on an image being displayed on a cathode-ray tube (CRT) monitor and a user wearing some form of shutter glasses. In order to display geometrically correct images, several systems include a head tracker that measures head pose in up to six degrees of freedom. A typical property of desktop VR systems is that users perform their task while sitting down. Therefore, and because of the limited display area, the extent of head motions is limited.

A special class of desktop VR systems is the class of systems that use a mirror to reflect the image of a CRT. The user of such a system can reach under the mirror, thus allowing for co-located interaction with the virtual world. The first such system was described by Schmandt [Sch83]. More recent examples include the work by Von Wiegand *et* 



*al.* [vWSS99] and the commercially available system from ReachIn Technologies AB [Rea].

Here we consider a mirror-based desktop VR system [MvL02], the Personal Space Station (PSS, Figure 1). The PSS is a desktop virtual reality system designed to work under office conditions. Interaction is performed directly in the 3D workspace by manipulating optically tracked input devices tagged with retro-reflective markers. The user's head pose is tracked in six degrees of freedom with the commercially available Logitech acoustical tracker.

## 1.1. Head Tracking

Tracking equipment providing six degrees of freedom at sufficient accuracy and performance can be expensive, both in terms of cost and compute time. Furthermore, tracking solutions often require some wired device physically attached to the user, which can be uncomfortable to wear and possibly hinder the user in freedom of movement. To leverage the latter problem, wireless (e.g. optical) tracking systems have been built that do not require the user to wear uncomfortable head gear [Rek95, Bir98, CSA00, MRCD02, MJvR03]. To our knowledge, no work addresses whether it is necessary to track the user's head pose in full six DOF for desktop VR systems.

This report investigates the validity of using a head tracker that tracks only lateral head motions in a desktop VR task. We expect the lateral translation component to be prevalent for desktop VR tasks, hence we expect that tracking only lateral head motions does not significantly affect user performance.

For this, we performed experiments to study head motions present in a desktop VR task. The results show the lateral translational component to be prevalent over the other translational components. Based on these results, we performed a user study to evaluate the validity of tracking only lateral head movements in a desktop VR task. The user study comprised a depth estimation task under monoscopic and stereoscopic viewing conditions, as well as no head-coupled perspective, laterally tracked head-coupled perspective and full 6 DOF tracked head-coupled perspective conditions.

The structure of this paper is as follows: Section 2 describes related work in this area. Section 3 briefly describes the set-up and results of our pilot study. The next section, Section 4, discusses the results of our user study. Finally, in Section 5 conclusions and suggestions for future work are given.

### 2. Related work

This paper is related to the work of Arthur *et al.* [ABW93] and Ware *et al.* [WAB93], who address user performance taking stereopsis and head coupled perspective as parameters, thus creating four viewpoint conditions: monoscopic

Cameras Mirror VFP

Figure 1: Schematic side view and a prototype of the Personal Space Station (PSS).

fixed, monoscopic head coupled, stereoscopic fixed and stereoscopic head coupled. They state that users commented on head coupled perspective giving a greater sense of depth than stereopsis alone. Furthermore, they show that the error rate for head coupled perspective in a tree tracing task was significantly lower than in the case of only stereopsis. These results are supported by the work of Rekimoto [Rek95], who performed the same experiment using a vision-based head tracking system.

Ware *et al.* [WF96] performed an experiment in which the size of an abstract three-dimensional graph that could be perceived under different viewing conditions was studied. The results showed that in the head-coupled stereo viewing case, the size of the graph that could be perceived by a subject increased by a factor of three. Furthermore, the experiment showed that head coupling alone was a better motion cue than stereoscopic viewing alone. In a second experiment, they showed that 3D motion cues were significantly better than cues provided by stereoscopic viewing in understanding the 3D abstract graph. The results of Ware's work are confirmed by the work of Arsenault *et al.* [AW00] who performed a study in which subjects had to perform a rapid hand movement task. The results show that head tracking has a positive effect on subject performance.

Faubert [Fau01] addresses stereoscopic viewing and head tracking in a theoretical framework and addresses how visual distortions affect depth perception from motion parallax.

Boritz and Booth [BB97] have performed a user study comprising a 3D point location task, in which they incorporate the same viewing conditions as in the aforementioned work. They state that head tracking had no appreciable effect on subject performance. The reasons they give for this are that motions are possibly hampered by the tracking device used, and a possible learning curve. On stereoscopic viewing they state it was superior to monoscopic viewing, which is in line with the work mentioned previously.

Fine and Jacobs [FJ99] compare motion and stereo cues in a depth perception task. Their work comprises an assessment

of several models for modeling the combination of motion, stereo and vergence cues. They conclude that the weighting of cues depends on task, viewing distance and noise model. In the work done by Yuan *et al.* [YSDSC00], depth perception from head motion parallax is shown not to be significantly affected when an artificially introduced time delay was kept below 265 milliseconds.

## 3. Pilot study

We have performed a pilot study to acquire data on the types of head motion present in desktop VR systems under a depth estimation task. The study comprises three different head tracking conditions being presented to the user: head coupled perspective disabled (condition NS), head coupled perspective enabled (condition HS) and head coupled perspective enabled with the possibility of objects occluding one another (condition FS). All conditions had stereoscopic viewing enabled.

Subjects were instructed to estimate the discrete depth level of a spherical reference point in a field of colored and numbered cubes (a *configuration*, Figure 2). To prevent fusion problems with cubes appearing close to each other on screen, each cube was given a different color.

The cubes were positioned on the surface of an invisible cylinder (Figure 3) with a radius of 5cm and a length of 20cm, the center laying at the center of the virtual workspace  $SW_0$  and rotated such that subjects looked down into the cylinder. Furthermore, the cylinder was divided into 10 discrete depth levels and 10 slices across. Each subject was presented a random permutation of the configuration, as well as random permutations of color and numbers for each configuration, so as to prevent users from making assumptions on depth level based on previous colors or numbers. Subjects were instructed to take as much time as needed for making a good estimation of the depth level of the reference point, during which head motions were recorded. Each subject was presented 25 configurations for each condition, thus giving a total of 75 configurations per subject.

#### 3.1. Pilot Study set-up

The pilot study was conducted in our mirror-based desktop VR system, the Personal Space Station (PSS). The PSS consists of a CRT monitor mounted in a frame, and a surface-reflecting mirror (see Figure 1). Users look into the mirror and perceive the image of the monitor to be behind the mirror. The display in the PSS is set to run at a resolution of  $1024 \times 768$  pixels at 120Hz. The center of the workspace is defined to lie 30cm above the table and to be aligned with the middle of the back of the display. The world coordinate system is defined by SW<sub>0</sub>, with the positive *x*-axis pointing to the user's right, the positive *y*-axis pointing upward and the positive *z*-axis pointing toward the user. The average viewing distance measured from the center of the eyes to the center

of the workspace is in the range of 50-70cm. The depth of the cubes varied from 40-60cm for the nearest and 60-80cm for the farthest, depending on the user. The distance measured from the point of the chin to the mirror was measured to be around 20cm, which we found not to limit the freedom of movement of the subjects.



**Figure 2:** Screen shot of application showing the reference point and the target cubes (depth range and object size increased for print and clarity)

Throughout the test, subjects wore synchronized NuVision 60GX shutter glasses (providing 60 images per second to each eye) with a Logitech acoustical head tracking device attached. The Logitech tracker reports 6 DOF pose data at up to 50 reports per second.

Eight subjects participated in the pilot study, all male, ranging in age from 22 to 36 with varying experience in VR environments.



Figure 3: Schematic side view of task

# 3.2. Results

Analysis of the translational data recorded during the pilot study shows the x ('lateral') component (for the user left-to-right) to be prevalent (Figure 4 shows an xy-plot of head position under tracking condition HS), with maximum lateral movement laying in the order of 20-30cm to each side. The standard deviation of translation in this direction is almost twice that of the other two components (Table 1).

Analysis of rotational components recorded during the study shows a maximum absolute rotation of about 4 degrees about the y and z axes. The maximum absolute rotation about the x axis was about 8 degrees. This can be explained by the fact that the center of the virtual scene is somewhat below and in front of the user. Therefore, under 'normal' viewing conditions, subjects will always look down toward the center of the scene.

	$\operatorname{std}_x$	stdy	$std_z$
NS: 0 DOF	7.709	2.363	4.736
HS: 6 DOF	7.440	3.238	3.827
FS: 6 DOF, occlusion	8.398	3.034	4.450

**Table 1:** standard deviation in centimeters for viewing conditions NS, HS and FS for the pilot study.

In addition to 6 DOF head pose measurements, we recorded the deviation of the depth level subjects mentioned from the correct depth level. The results showed that, on an overall perspective, subjects were able to successfully complete the given task.



**Figure 4:** *xy-plot of translation data recorded during the pilot study under head-coupled viewing conditions (condition* HS).

## 4. User Study

Based on the head movement data from the pilot study, which showed lateral translation to be prevalent, a new experiment was performed. The new experiment comprises a modified version of the task performed in the pilot study, as well as incorporating different viewing conditions. The following sections describe the user study in more detail and discuss the results.

## 4.1. Set up

The testing application was modified, in order to gain a more continuous scale of user performance (as opposed to the discrete scale in the pilot study). To increase the chance of inter-target occlusion, the virtual scene presented was scaled down: we decreased the size of the invisible cylinder (radius was decreased to 2cm, length was decreased to 15cm). Furthermore, the target cubes and reference point were scaled down, the target cubes being 1.5cm in size, and the reference point having a radius of 2mm. We found that modifying the task in this manner made the task difficult enough in the most information-rich situations, while still being doable in the less-rich situations.

The task for each subject in this test was to manually place the reference point at a depth level corresponding to a numbered cube. Positioning of the target cube was performed by moving a standard mouse forward and backward (movements sidewards had no effect), thus allowing the reference point to move along a fixed depth-axis through the center of the cylinder (schematically shown in Figure 3).

As in the pilot study, each subject was allowed to take as much time as needed for placing the reference point at the instructed level. Subjects were asked to indicate when they thought the reference point was at the correct depth. If a subject indicated the point was at the correct level, a new configuration was presented and the subject was given the number of a new target cube.

Before a new subject started with the actual test, he or she was given time to practice and to get used to viewing in stereo. During this practice period, a number of example configurations were shown during which the task was explained. When the subject indicated to be ready for the actual test, the application was re-started and the test began.

# 4.2. Conditions

In our user study, we used different viewing conditions to study the effect of tracking only lateral translation: no head-coupled perspective (condition NS), lateral-only head-coupled perspective (condition LS) and full head-coupled perspective (condition HS). In addition to the above (stereo-scopic) conditions, monoscopic viewing was introduced as a parameter (denoted by the *M*-suffixes), bringing the total number of viewing conditions to 6 for this study.

During the user study, the order of conditions and configurations was randomly permuted. Each subject was presented a sequence of 25 configurations for each condition, giving a total of 150 configurations per subject. After each viewing condition, subjects were allowed to remove the glasses and relax the eyes if they wished. During this period, subjects were asked if they had any comments on the test run just performed.

The subjects participating in the user study were all computer scientists with little experience in working with virtual environments. Of the eight subjects, two were female. Subjects ranged in age from 23 to 45 years old. Three of the subjects wore glasses. One subject of the user study group had previously participated in the depth estimation task for the pilot study.

During the study, we recorded user performance by measuring the deviation in distance of the reference point to the instructed depth level along the axis of movement of the reference point. In addition, we recorded six degrees of freedom head pose data, the time it took a subject to make the final estimate for a configuration, and any remarks subjects had during the test run.

## 4.3. Results

The following sections discuss the results from the user study. First, we discuss head motion data gathered from the user study. After this, we address user performance under the different viewing conditions of the user study and discuss the results.

# 4.3.1. Head motion

Analysis of head motion data from the user study shows no significant difference in head movements in the *xy*-plan: under all conditions, a clear arc-shaped figure can be discerned (similar to Figure 4). An explanation for this arc-shaped figure can be found in the relative rigidity of the human spine.

#### 4.3.2. User performance

During the experiment, the distance of the position of the reference point positioned by a subject to the instructed depth level was recorded. The mean absolute error in subjects' estimates are shown in Table 2 (all values in centimeters).

	NS	LS	HS	NM	LM	HM	
1	1.550	0.266	0.246	6.354	0.702	0.586	
2	0.378	0.266	0.238	6.150	0.790	0.482	
3	0.554	0.714	0.298	35.138	1.870	2.218	
4	0.286	0.410	0.394	20.002	16.798	8.886	
5	0.378	0.294	0.378	21.482	8.526	4.046	
6	1.090	0.306	0.354	4.810	0.662	0.522	
7	0.370	0.274	0.262	9.658	14.700	3.770	
8	0.314	0.214	0.262	6.106	1.762	0.834	
μ	0.615	0.343	0.304	13.712	5.726	2.668	
Table 2. Maan abaaluta amana in aantimatana							

 Table 2: Mean absolute errors in centimeters.

An analysis of variance (ANOVA) within the stereoscopic viewing conditions NS, LS and HS, shows that there is no significant difference in user performance between the lateral tracking and full tracking conditions (LS and HS, respectively, p < 0.001 at the 0.05 significance level). However, user performance under both these conditions was significantly better than under the non-tracked (NS) condition. Analyzing variance within the monoscopic viewing group showed no significant differences in performance. Subjecting all six conditions to an ANOVA shows that all conditions, except the LS and HS-conditions, differed significantly. Figure 5 shows a box-whisker plot for all six viewing conditions; Figure 6 shows a box-whisker plot for only the stereoscopic viewing conditions.

In addition to errors in depth estimation, we recorded the



Figure 5: Box-whisker plot of subject performance (absolute error in cm), see Table 2. The grayed box is the inter-quartile range, showing the median of the data as a horizontal black line within the box. Outliers in the data set are drawn as black dots. Note that for image scale, a number of outliers have not been drawn. They are, however, used in the calculations.



Figure 6: Box-whisker plot of subject performance (absolute error in cm) for the stereoscopic viewing conditions. The vertical axis range has been set to 1.5 cm (resulting in some outliers not being drawn) for clarity

time it took each subject to place the reference point at the instructed depth. Analysis of these data showed no significant differences in the time subjects needed to perform the task throughout the viewing conditions.

#### 4.3.3. Subjective evaluation

Not surprisingly, all subjects commented on the monoscopic, non-head coupled perspective condition NM on as being the most difficult. Under this condition, the only depth cues available were relative movement of the reference point toward and away from the user, and changes in relative size of the reference point when moving the point. A number of subjects mentioned they based their judgment of depth in this case on the relative size of the reference point. This often resulted in placing the reference point in front of the field of target cubes, which might be explained by the smaller relative size of the reference point.

Subjects often noticed that something had changed when a stereoscopic condition was followed by a monoscopic condition, but were not always able to exactly tell what had changed.

Most subjects, when asked for differences between conditions, commented that they found both the laterally tracked and fully tracked monoscopic conditions (*LM* and *HM*, respectively) equally difficult.

One user commented that a different input device, such as a lever would be preferable for placing the reference point at the specified depth. The test results showed that subjects did not have any problems with positioning the reference point with the mouse.

The subject that had participated in the pilot study commented on the second experiment being tougher than the first and that he felt that head movement was of greater importance than during the first experiment in which he participated.

# 4.4. Discussion

Our results show that user performance was not significantly affected when incorporating only lateral head motion under stereoscopic viewing conditions. Under the monoscopic viewing conditions, analysis of the variance showed all conditions to be significantly different. Even though the difference between the laterally tracked and the fully tracked conditions is significant in these conditions (monoscopic viewing), the difference in user performance comparing the tracked conditions with the non-tracked condition is relatively large.

## 5. Conclusion

In this paper, we addressed the validity of using a single degree of freedom head tracking device, tracking only lateral translation components for a desktop virtual reality system. Our results show that incorporating only lateral head motion does not significantly adversely affect user performance in a desktop VR task. The results furthermore show that both the laterally tracked and fully tracked cases result in significantly better user performance than the non-tracked case.

Furthermore, we conclude that user performance was not significantly affected by (small) distortions of visual space caused by tracking only lateral head motions. Systems employing a full six degree of freedom head tracker can therefore use the 'extra' measurements to decrease distortions and providing the user with a possibly more comfortable view (although no subject commented on anything but the *NM*-condition being uncomfortable or unnatural). It is, however, not to be expected that the addition of extra degrees of freedom will significantly improve user performance.

Our results show a large difference in user performance between the monoscopic and stereoscopic viewing conditions. Furthermore, our results show that the increase in user performance is largest in the monoscopic case when head tracking is employed, compared to the non-tracked case. Also, the increase in user performance is larger when going from the *LM*-condition to the *HM*-condition, than when going from the *NM*-condition to the *LM*-condition. From this, we conclude that stereopsis is an important depth cue for our type of task, which is in line with previous work. Furthermore, we conclude that tracking head pose in full six degrees of freedom is more important under monoscopic viewing than under stereoscopic viewing conditions.

Several researchers have shown head-coupled stereoscopic viewing to be an important depth cue, something that is also true for desktop VR systems. Desktop VR systems are found in different forms ranging from full-fledged stereoscopic viewing systems to a simple monitor with a web cam for tracking the user's head. Although we believe our work is relevant for the whole class of desktop stereoscopic VR systems, future work is needed to assess the effect on different system types (e.g. in augmented reality, where small distortions are much more noticeable than in the case of virtual reality) and different types of tasks.

#### References

- [ABW93] ARTHUR K. W., BOOTH K. S., WARE C.: Evaluating 3d task performance for fish tank virtual worlds. ACM Transactions on Information Systems (TOIS) 11, 3 (1993), 239–265.
- [AW00] ARSENAULT R., WARE C.: Eye-hand coordination with force feedback. In Proceedings of the SIGCHI conference on Human factors in computing systems (2000), ACM Press, pp. 408–414.
- [BB97] BORITZ J., BOOTH K. S.: A study of interactive 3d point location in a computer simulated

virtual environment. In *Proceedings of the ACM symposium on Virtual reality software and technology* (1997), ACM Press, pp. 181–187.

- [Bir98] BIRCHFIELD S.: Elliptical head tracking using intensity gradients and color histograms.
   In *IEEE Conference on Computer Vision and Pattern Recognition* (1998 1998), IEEE.
- [CNSD93] CRUZ-NEIRA C., SANDIN D. J., DEFANTI T. A.: Surround-screen projection-based virtual reality: The design and implementation of the cave. In *Proceedings of SIGGRAPH 93* (August 1993), pp. 135–142.
- [CSA00] CASCIA M. L., SCLAROFF S., ATHITSOS V.: Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3d models. *IEEE Transactions* on Pattern Analysis and Machine Intelligence 22, 4 (2000), 322–336.
- [Dee92] DEERING M.: High resolution virtual reality. In Proceedings of the 19th annual conference on Computer graphics and interactive techniques (1992), ACM Press, pp. 195–202.
- [Fau01] FAUBERT J.: Motion parallax, stereoscopy, and the perception of depth: Practical and theoretical issues. In *Three-dimensional video and display: devices and systems* (2001), Javidi B., Okano F., (Eds.), vol. CR76, The International Society for Optical Engineering (SPIE), pp. 168–191.
- [FJ99] FINE I., JACOBS R. A.: Modeling the combination of motion, stereo and vergence angle cues to visual depth. *Neural Computation 11* (1999), 1297–1330.
- [McK92] MCKENNA M.: Interactive viewpoint control and three-dimensional operations. In Proceedings of the 1992 symposium on Interactive 3D graphics (1992), ACM Press, pp. 53–56.
- [MJvR03] MULDER J. D., JANSEN J., VAN RHIJN A.: An affordable optical head tracking system for desktop VR/AR systems. In Proceedings of the workshop on Virtual environments 2003 (2003), ACM Press, pp. 215–223.
- [MRCD02] MORENCY L.-P., RAHIMI A., CHECKA N., DARRELL T.: Fast stereo-based head tracking for interactive environments. In *Proceedings* of Conference on Automatic Face and Gesture Recognition (2002).
- [MvL02] MULDER J. D., VAN LIERE R.: The personal space station: Bringing interaction within reach. In *Proceedings of VRIC 2002;*

4th Virtual Reality International Conference (2002), pp. 73–81.

- [Pfa96] PFAUTZ J. D.: Distortion of Depth Perception in a Virtual Environment Application. Master's thesis, Massachusetts Institue of Technology, 1996.
- [Pfa00] PFAUTZ J. D.: Depth Perception in Computer Graphics. PhD thesis, Trinity College, University of Cambridge, May 2000.
- [Rea] REACHIN TECHNOLOGIES AB: Reachin display. http://www.reachin.se.
- [Rek95] REKIMOTO J.: A vision-based head tracker for fish tank virtual reality – vr without head gear. In Proceedings of IEEE Virtual Reality Annual International Symposium (VRAIS '95) (1995), pp. 94–100.
- [Sch83] SCHMANDT C.: Spatial input/display correspondence in a stereoscopic computer graphic work station. In *Proceedings of the 10th annual conference on Computer graphics and interactive techniques* (1983), ACM Press, pp. 253–261.
- [vWSS99] VON WIEGAND T. E., SCHLOERB D. W., SACHTLER W.: Virtual workbench near-field virtual environment system with applications. *Presence: Teleoperators and Virtual Environments* 8, 5 (October 1999), 492–519.
- [WAB93] WARE C., ARTHUR K., BOOTH K. S.: Fish tank virtual reality. In Proceedings of the SIGCHI conference on Human factors in computing systems (1993), ACM Press, pp. 37–42.
- [WF96] WARE C., FRANCK G.: Evaluating stereo and motion cues for visualizing information nets in three dimensions. ACM Transactions on Graphics 15, 2 (1996), 121–139.
- [WHR99] WARTELL Z., HODGES L. F., RIBARSKY W.: Balancing fusion, image depth and distortion in stereoscopic head-tracked displays. In Proceedings of the 26th annual conference on Computer graphics and interactive techniques (1999), ACM Press/Addison-Wesley Publishing Co., pp. 351–358.
- [YSDSC00] YUAN H., SACHTLER W. L., DURLACH N., SHINN-CUNNINGHAM B.: Effects of time delay on depth perception via head-motion parallax in virtual environment systems. *Presence: Teleoperators & Virtual Environments 9*, 6 (December 2000), 638–647.