The Personal Space Station: Bringing Interaction Within Reach

Jurriaan D. Mulder and Robert van Liere
Center for Mathematics and Computer Science, CWI
Amsterdam, the Netherlands
{mullie,robertl}@cwi.nl

Abstract

Near-field virtual reality allows users to interact with virtual objects within arm's reach of the user. Environments for near-field VR are well suited for direct precise interaction by taking advantage of the user's hand-eye co-ordination.

We discuss the design and initial experience of a near-field virtual environment, the Personal Space Station (PSS). In this system, all interactive 3D tasks are realized directly with the hands or by using task specific graspable input devices. The primary motivations for building the system are to provide an environment that can be used under normal office working conditions, that allows for direct natural interaction, and that is low cost.

The PSS consists of a mirror in which stereoscopic images are reflected. The user reaches under the mirror to interact with the virtual world. The principle differences between the PSS and other near-field systems are that interaction is based on optical tracking, the system is configurable to adapt to various application requirements, and the PSS is a low-cost, complete desktop VR system. Two cameras are used to track the space in which the interaction takes place. Robust and low latency optical tracking paves the way for versatile and graspable 3D interfaces supporting direct interaction.

1 Introduction

In this paper we describe the Personal Space Station (PSS), a near-field virtual environment that addresses the issues of direct interaction, ergonomics, and costs. The PSS consists of a mirror in which a stereoscopic image is reflected. The user reaches under the mirror to interact with the virtual objects directly with his hands or by using graspable, task specific input devices. Cameras are used to track interaction. A prototype PSS is shown in Figure 1.

The motivation for building the PSS is threefold. First, to provide an environment for 3D applications based on wireless, direct and natural interaction; the PSS uses optical tracking for interaction. Second, to provide an environment that can be used in normal office working conditions; the PSS is designed to fit on a user's desk and can be used under normal lighting conditions while the user is comfortably seated. Third, to provide a versatile design that allows a low-cost system to be built; the PSS is constructed using only off-the-shelf commodity components.

Until now, interaction in virtual environments has mostly been realized with generic wire-based 6 degrees-of-freedom (DOF) input devices, such as a wand or stylus equipped with a magnetic tracker. Interaction techniques that use these devices are often indirect, difficult to use, and lack precision which can result in loss of task performance, discomfort, and user fatigue. Near-field interactive tasks should be done in a more direct and natural way, using either task specific input devices or directly with the hands. Fitzmaurice for example, has had great success with graspable user interfaces [1]. Recent research indicates that hand-image co-location can increase task performance significantly, particularly when orientation is involved [2]. However, when used in back-projected display systems near-field interaction is problematic: hands and other body parts block the display resulting in occlusion of the image, and the physical projection surface prohibits direct interaction with the virtual world behind the surface. Head-mounted display systems (HMDs) would be better suited for near-field direct interaction. However, HMDs still lack sufficient display quality, or they become very expensive [3].

A different approach for near-field virtual environments is to use a mirror in which a stereoscopic image of the virtual world is reflected from a dis-

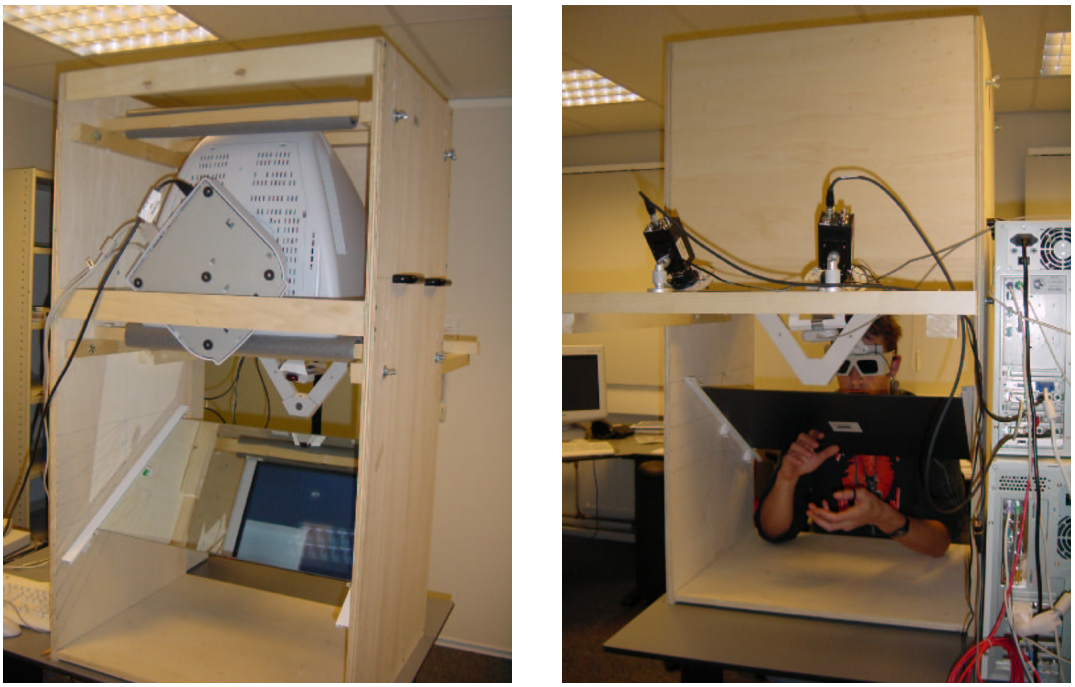


Figure 1: The Personal Space Station prototype. Left: front view. Right: rear view.

play surface. A user can reach under the mirror into the virtual world to interact. The principle advantage of using a mirror is that the 3D space in which interaction is performed is superimposed over the visual space without obscuring the image. In this way, a 3D interface can take advantage of the user's unique hand-eye co-ordination to stimulate various sensory-motor cues. Indeed, such reach-in environments are often referred to as 'dexterous VR', in order to emphasize on the skillful and competent usage of hands.

In this paper we present the design and initial experience of the prototype Personal Space Station, a mirror-based, near-field desktop VR system. The next Section discusses related work and positions the PSS against other environments. In Section 3 we describe the PSS in more detail, followed by some examples of 3D interaction in the system (Section 4). In Section 5, we discuss initial user experience with the prototype PSS and a number of design trade-offs.

2 Related Work

The first mirror-based environment was proposed by Schmandt in 1983 [4]. The system used a half-silvered mirror with a standard CRT monitor, and a 6 DOF magnetic tracking device with a stylus for positional input. Schmandt's goal was to allow "a style of interaction in which spatial correspondence between input and output devices could be maintained". Although advanced for its time, the system suffered from magnetic interference with the tracker caused by the CRT. Poston et al. [5] and Wiegand et al. [3] propose similar systems but equipped with respectively a mechanical tracker and a Phantom for 3D input and force-feedback. The Swedish company ReachIn [6] sells a reach-in device which also uses a Phantom for 3D input and force-feedback. In all these systems, the user's head position is not tracked. Poston et al. for instance, claim that head tracking introduces too much latency for precise interactive work. The PSS differs from these systems in two ways: first, since we require that images are correctly projected, in our system the user is head tracked. Second, our system uses optical tracking for all 3D interactive tasks.

It would appear that a correct perspective view is mandatory in order for normal hand-eye coordination skills to be applied. However, recent user studies have not been conclusive on whether head tracking increases the performance of 3D interactive tasks. For example, Boritz and Booth report that head tracking has no appreciable effect on the performance of 3D positioning tasks [7]. On the other

hand, Arsenault and Ware do find that head tracking has a positive effect on the performance of a rapid hand movement task [8]. Although head tracking accounts for additional complexity and cost of the system, we postulate that a correct perspective view is necessary. Our experience is that the depth cues gained from motion parallax are substantial for the (bio-medical) applications we are studying. Moreover, since our future plans include using a half-silvered mirror to implement augmented reality applications, we believe that head tracking is necessary to maintain the spatial correspondence between the virtual and real world.

In our system, retro-reflective markers under infra-red (IR) lighting conditions are used to track objects located under the mirror. Advantages of optical tracking are that it allows for wireless input, it is less susceptible to noise, and it allows for many objects to be tracked simultaneously. The usage of retro-reflective markers for virtual reality is not new. For example, Ribo et al. [9] and Dorfmüller [10] report on similar tracking systems using retro-reflective markers for wand-like interaction and head tracking on a projection table. Although no quantitative measurements are given, Ribo et al. claim "very good spatial accuracy in all 6 DOF". The PSS differs from these approaches in that, due to compact space constraints, we require a much closer range image acquisition and object recognition.

A system similar to the Personal Space Station is the Virtual Hand Lab (VHL), developed at Simon Fraser University [11]. However, the VHL differs from the PSS in a number of ways. The VHL is intended to be an experimental test-bed to examine human perception and motor performance when complex tasks are executed using the hands. The PSS is designed as a complete system which can be used in multiple application areas, including scientific visualization, training, and entertainment. Also, the costs of the PSS are more than an order lower than the VHL, bringing the PSS within reach of every end-user. Further, the working volume of the PSS can be easily configured for different application requirements.

3 The PSS Prototype

3.1 Apparatus

The apparatus consists of three components: a graphics engine, a tracking engine, and a wooden chassis to mount the monitor, the mirror, and the cameras (see Figure 1). The construction of the $65 \times 125 \times 55$ cm chassis allows for easy experimentation with different workspace configurations by adjusting the mount points of the mirror, the monitor, and the cameras used for tracking. The graphics engine is a standard PC equipped with an ATI FireGL3 graphics board and a high-resolution 19 inch CRT monitor. The display resolution is set to 1024x768 @ 120hz. The tracking engine is a PC equipped with two Leutron Vision PictPort H4D dual channel frame grabbers and two Leutron Vision LV-7500 progressive scan CCD-cameras. Two Computar H0612FI lenses with a focal length of 6 mm and an F number of 1.2 are fitted to the cameras.

The total hardware costs of the prototype PSS are approximately 13 kEuro. Although still expensive, the costs are substantially lower than for instance a large back projected display environment. The price of one single high quality BARCO projector is significantly higher than the complete prototype.

The prototype PSS runs under the Linux operating system. The FireGL is currently the only graphics board with Linux device drivers that support quad-buffered stereo. PVR, an in-house toolkit for portable virtual reality applications [12], is used for software development.

3.2 Design

The basic design of the system is diagrammed in Figure 2. The design distinguishes between three spaces: the *visual space* (defined as the virtual space that the user can visually perceive), the *interaction space* (the area in which the user performs 3D interaction), the *tracking space* (the area covered by the cameras).

Applications may require different workspace configurations. Several parameters influence the configuration: the position, orientation, and size of the CRT monitor, the position and orientation of the mirror, the intrinsic and extrinsic parameters of the cameras, and the position of the user with respect to the chassis. The goal is to choose these parameters such that the visual space, the interaction space, and the tracking space coincide.

Visual Space The mirror reflects the display surface of the CRT monitor into a *virtual focus plane* in front of the user. Due to accommodation and convergence conflicts, the useful depth range of the visual space is limited. This depth range should not exceed +/- 10 centimeters around the focus

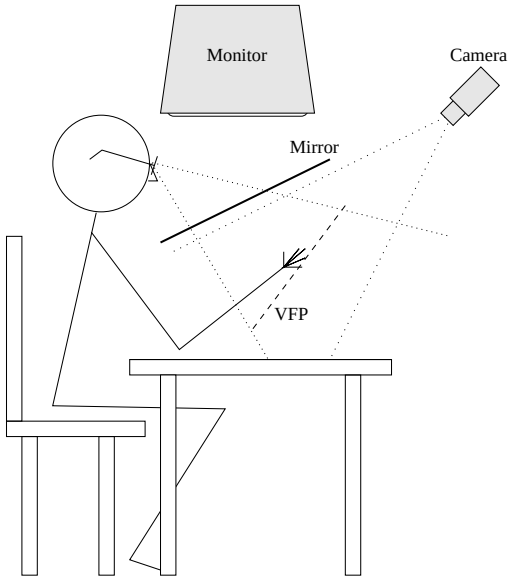


Figure 2: Schematic side view of the Personal Space Station. Indicated are the monitor, the mirror, the cameras, and the configurable virtual focus plane (VFP). The stippled lines indicate the visual space and tracking space. The head tracked user is comfortably seated and interaction is direct.

plane. Virtual objects drawn outside the depth range may cause visual discomforts such as double vision (when the user is unable to fuse the stereoscopic images). Depending on the application requirements, the CRT monitor and the mirror can be mounted at a different positions and orientations in the chassis. Changing these positions and orientations will result in a different position and orientation of the virtual focus plane with respect to the chassis.

A Logitech 6 DOF acoustic head tracker is used for head tracking. The ultrasound emitter is mounted in the chassis above the mirror and the receiver is mounted on the user's shutter glasses. The active area of the head tracker is defined as a 100-degree cone that extends approximately five feet from the transmitter. The published resolution of the tracker is 1/250 of an inch along the X,Y, and Z axes and 1/10 of a degree for the pitch, yaw and roll rotations. The tracker can generate 50 reports per second and the reported minimal latency (without filtering) is 30 ms.

Interaction Space The interaction space is restricted to the area that the user can reach with his hands or with the input devices. The design goal is to position the interaction workspace such that it encloses the visual workspace and 3D interaction can be realized comfortably, i.e. the user is seated behind a desk, his elbows are rested on the desk top, and he should not need to over-reach into the virtual world to perform 3D interaction. Important parameters in this respect are the position and height of the chair and table in combination with the user's physical characteristics such as his upper and lower arm length. Furthermore, the PSS should also allow for conventional (non 3D) interaction devices to be used, such as a keyboard or dials.

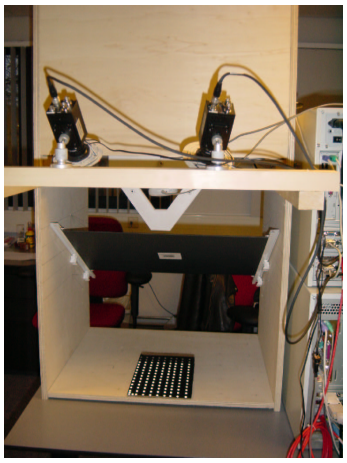
Tracking Space A camera's field of view is determined by its extrinsic parameters (position and orientation) and intrinsic parameters (internal geometry and optical characteristics). The tracking space is defined as the intersection volume of both cameras' field of view. The tracking space is illuminated by rings of IR leds mounted closely around the camera lenses. IR-pass filters in front of the camera lenses are used to cut-off the light below a chosen wavelength. Retro-reflective markers are applied to all objects to be tracked. IR light from the leds is reflected by the markers into the lens such that, after thresholding, blobs of white pixels occur in the acquired image.

For the configuration of the tracking space, it is our desire to meet two requirements: we wish to mount the cameras as close as possible to the workspace (to keep the PSS compact) while acquiring a tracking space that encloses the 3D interaction space. Several trade-offs come into play. For example, to position the cameras very close to the workspace, lenses with small focal lengths have to be chosen. Such lenses however, cause significant distortions in the acquired images that have to be corrected. Furthermore, IR light reflected by normal (diffuse reflective) objects that are close to the lens can cause undesired blobs in the image, making it more difficult to track the retro-reflective markers robustly.

The steps performed in reconstructing a 3D position from the marker reflections in the images consists of 2D blob position detection, correction, rectification, corresponding, and 3D re-projection. These steps are well known from literature, see for instance [13]. To facilitate the correspondence problem, markers of different sizes and shapes can be used. In addition, markers can be placed in specific patterns.

For camera calibration and parameter estimation, we use a method as developed by Zhang [14]. The method computes 6 extrinsic camera parame-

ters (position and orientation) and 8 intrinsic (focal length, aspect ratio, image center, and two radial and two tangential distortion coefficients). Figure 3 tabulates an initial measurement of the accuracy of the optical tracker. For this we placed 108 markers (9 rows, 12 columns) in a rectangular pattern. The rectangular pattern is placed at three different heights in the workspace. The accuracy is defined as the difference between the real and computed positions. The image in Figure 3 shows the pattern placed at a height of 3 centimeters. The table gives the minimum, maximum, and average error found at each height.



Height	Min	Max	Avg
3.0 cm	4 mm	7 mm	6 mm
14.0 cm	1 mm	4 mm	2 mm
24.2 cm	2 mm	6 mm	4 mm

Figure 3: Accuracy measurements: the image shows the rectangular pattern of 108 markers in the workspace at height = 3 cm. The table gives the minimum, maximum and average accuracy of markers at three different heights. Accuracy is defined as the difference in millimeters of the marker position in the real world and the reconstructed 3D position.

The table shows that 3D positions are computed with an accuracy between 1 and 7 millimeters. The numbers given in the table are from initial calibration and measurement procedures. Although we are encouraged by these results, we believe that the accuracy can be maximized when we improve our calibration and measurement procedures.

4 Interaction Examples

To demonstrate the capabilities of the prototype PSS, a simple interactive molecule viewer has been implemented. The molecule viewer uses four input devices and interaction techniques (see Figure 4).

The *thimble* device is used for pointing and selecting atoms. Device feedback is provided by drawing a cone and highlighting the selected atom. The position and direction of the thimble is tracked with two reflective markers. Atoms intersecting the tip of the thimble are selected. The *cutting plane* device is constructed as a round cardboard plane mounted on a pencil. It is used to cut a molecular surface. The position and direction of the cutting-plane device is tracked with two reflective markers on the pencil. Feedback is drawn as a transparent cutting plane at the same position of the round cardboard plane. Finally, the *ruler* device is used for distance measurements in the molecule. The ruler device is constructed as a sliding piece of cardboard mounted on a plastic ruler. The ruler is tracked with two reflective markers, one on the tip of the plastic ruler and one on the sliding cardboard. Measurements are made snapping the tip of the ruler on an atom and sliding the cardboard to other atoms. Device feedback is given by drawing a segmented cylinder. The length of each segment is molecule specific, but typically represents a few nano-meters.

Two handed input is possible by, for example, using a *cube* device in the non-dominant hand simultaneously with the cutting plane device in the dominant hand (Figure 5). The cube device is used to position and orient the molecule. Device feedback is provided by drawing a small coordinate system to indicate the orientation of the molecule. Patterns of 4 reflective markers are tracked which are used to compute the position and orientation of the cube. The right image of Figure 5 shows an example of the detected blobs from one of the cameras.

The tracking performance in these techniques is adequate. The dual cameras and frame grabbers are able to grab 60 PAL images per second. Blob detection, correction, rectification, correspondence, and 3D re-projection can be done in near real time. For each device, the tracking engine sends more than 50 reports per second to the graphics engine. The maximum latency of each device is between 25 and 50 milliseconds.

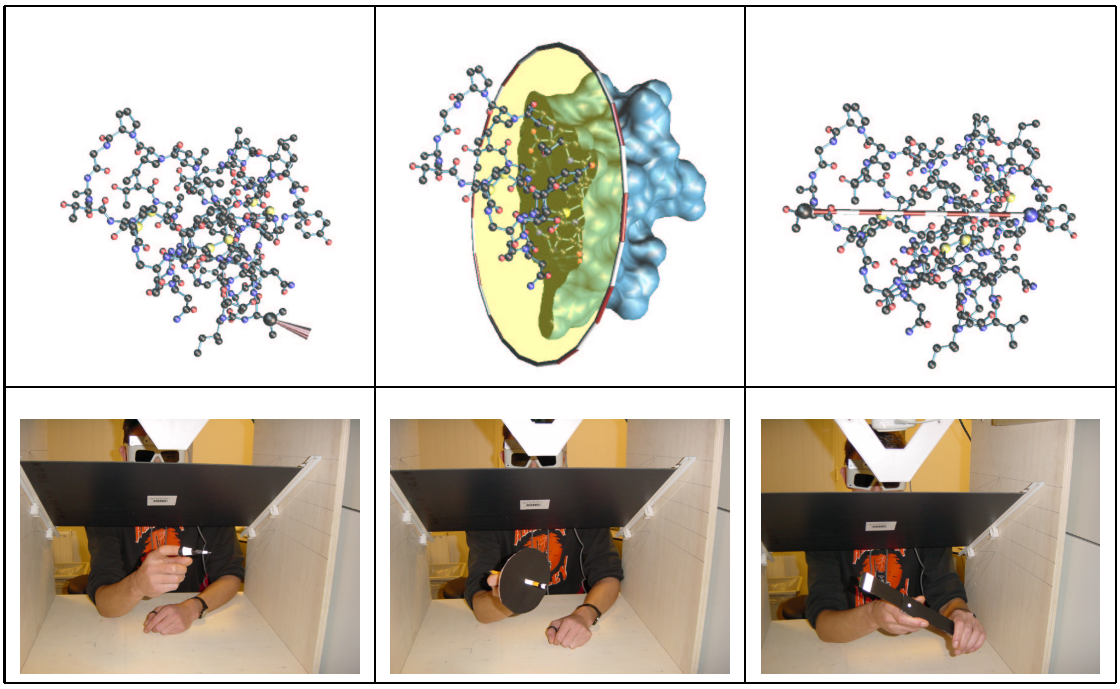


Figure 4: Some low-cost, wireless input devices. Bottom row from left to right: a thimble used for atom selection, a cutting plane device to position and orient a cutting plane, and a ruler for distance measurements. The top row images show the corresponding device feedback.

5 Discussion

It is our experience that in near-field VR, users are being 'attracted' to the objects seen; as soon as an object comes within arm's reach the natural reaction of the user is often to reach out and try to manipulate the object. It is this notion of hand-eye co-ordination that we want to exploit in our system: all 3D interactive tasks are realized directly with the hands or by using simple task specific input devices.

We have demonstrated our system with several example applications to many users, both experienced and people novel to VR. Through informal observations and from user comments we found that interaction indeed 'comes natural'. The combination of simple, wire-less interaction devices, hand-eye co-ordination, and visual feedback relieves the user of having to reason about the desired interaction to be performed.

The motivations for designing the PSS were to address the issues of natural interaction, ergonomics, and costs. To enable interaction, we have chosen for optical tracking in which each object to be tracked is equipped with retro-reflective markers. An inherent problem to this approach is that of marker occlusion, which will result in a (temporary) loss of markers in the camera images. Various methods have been proposed by the computer vision community that address this problem [15]. For example, to minimize the problem of occlusion, additional markers can be placed on an object or additional cameras can be placed in the scene. An other approach is that of model based interaction in which additional information about an object and its markers is known at all times. For example, in addition to a position, a marker may have a velocity and acceleration. In this way, higher-level decisions about object identification and positioning can be made.

With respect to ergonomics, it has been our desire to make the design 'user and office-friendly'. This includes keeping the system compact and portable, being able to use it in normal office conditions, and being able to use the system comfortably while being seated with the elbows rested on the desk top. A few remarks to be made from experience are that:

- The Logitech acoustic head tracker causes discomfort because the receiver is mounted on the stereo shutter glasses which makes it too bulky and heavy. Also, the active area of the tracker at close range is too small. A different approach for head tracking would be to place three reflective markers on the shutter glasses

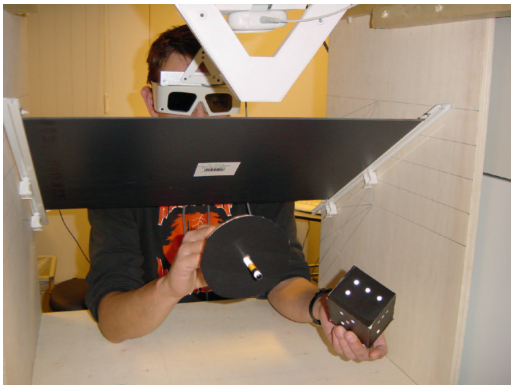


Figure 5: Two handed input: the cube is used to position and orient the molecule and the cutting plane device is used to cut the surface. The right image is an example snapshot of the patterns captured from one of the cameras.

and use optical tracking. However, this approach would require the tracking space to enclose the user's head or the use of additional cameras.

- The display suffers from ghosting: beside the correct image, the left and right eyes also perceive a (low intensity) image intended for the other eye. This may be caused by inadequate closure of the shutter glasses or by too much after-glow of the monitor. It looks awkward and can cause fusion problems. Furthermore, the display rate of 120 Hz interferes with normal office neon lights, which causes an annoying flicker in the peripheral vision.

With respect to costs, the hardware of the prototype PSS amounts to approximately 13 kEuro. More than half of this is due to the optical tracking hardware. The prototype uses two Leutron LV-7500 progressive scan CCD-cameras. The frame grabbers in combination with these cameras provide full stereo images in PAL resolution at 60 frames a second. A significantly cheaper approach would be to use interlaced CCD-cameras. These cameras can acquire images in PAL resolution at 30 frames a second using only a single frame grabber. However, each image consists of two interlaced images that may be slightly displaced. In the near future, we want to investigate the effect of using interlaced cameras on the overall accuracy and performance of the tracking system. In general, feature detection on slightly displaced images may result in less accurate point reconstruction. However, smoothing and prediction filters might be used to accommodate for loss of accuracy and performance.

6 Conclusion

In this paper we describe the Personal Space Station (PSS), a virtual environment for applications that employ precise near-field interaction. The goal of the PSS is to allow the user to interact directly with a virtual world. Special emphasis of the design has been placed on the ergonomics and costs of the system. The main differences between the PSS and other near-field systems is that interaction is based on optical tracking, the system is configurable to adapt to various application requirements, and the PSS is a low-cost, complete desktop VR system.

Experience with near-field VR systems has shown us that users are attracted to virtual objects within reach. They are eager to grasp and manipulate the virtual objects with their hands. We believe that by enabling users to interact with these objects in such a direct and natural way, and thereby exploiting the notion of hand-eye co-ordination, the feeling of presence will be enhanced and interaction will be more transparent and accurate.

In the near future, we want to formally investigate these findings with user studies. Furthermore, we will continue to improve the design and implementation of the PSS according to the experiences obtained from these studies and from the usage of the PSS in bio-medical applications. Additional enhancements include the development of a more compact and lightweight chassis, the mounting of a half-silvered mirror to study graspable 3D interfaces in augmented reality, and the development of higher level model-based interaction methods to overcome the occlusion problem inherent to optical tracking.

References

[1] G.W. Fitzmaurice. *Graspable User Interfaces*. 1996. Ph.D. Thesis, Dept. of Computer Science, Univ. of Toronto.

[2] C. Ware and J. Rose. Rotating virtual objects with real handles. *ACM Transactions on CHI*, 6(2):162–180, 1999.

[3] T. von Wiegand, D. Schloerb, and W. Sachtler. Virtual workbench: Near field virtual environment system with applications. *Presence*, 8(5):492–519, 1999.

[4] C. Schmandt. Spatial input/display correspondence in a stereoscopic computer graphic work station. *Computer Graphics*, 17(3):253–261, 1983.

[5] T. Poston and L. Serra. Dextrous virtual work. *CACM*, 39(5):37–45, 1996.

[6] http://www.reachin.se.

[7] J. Boritz and K. S. Booth. A study of interactive 3D point location in a computer simulated virtual environment. In *ACM Symposium on Virtual Reality Software and Technology*, pages 181–187, 1997.

[8] R. Arsenault and C. Ware. Eye-hand coordination with force feedback. In *CHI 2000 Conference Proceedings*, pages 408–414, 2000.

[9] M. Ribo, A. Pinz, and A. Fuhrmann. A new optical tracking system for virtual and augmented reality applications. In *Proceedings of the IEEE Instrumentation and Measurement Technical Conference*, pages 1932–1936, 2001.

[10] K. Dorfmüller. Robust tracking for augmented reality using retroreflective markers. *Computers and Graphics*, 23(6):795–800, 1999.

[11] V.S. Summers, K.S. Booth, T. Calvert, E. Graham, and C.L. MacKenzie. Calibration for augmented reality experimental testbeds. In *Proceedings of the 1999 Symposium on Interactive 3D Graphics*, pages 155–162, 1999.

[12] R. van Liere and J.D. Mulder. PVR - an architecture for portable VR applications. In M. Gervautz, A. Hildebrand, and D. Schmalstieg, editors, *Virtual Environments '99, Proceedings of the Virtual Environments Conference & 5th Eurographics Workshop*, pages 125–135. Springer Verlag, 1999.

[13] S. Barnard and M. Fischler. Computational stereo. *ACM Computing Surveys*, 14(4):553–572, 1982.

[14] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.

[15] G. Klinker, K. Ahlers, D. Breen, P.-Y. Chevalier, C. Crampton, D. Greer, D. Koller, A. Kramer, E. Rose, M. Tuceryan, and R. Whitaker. Confluence of computer vision and interactive graphics for augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4):433–451, 1997.