CCD-Camera Based Optical Tracking for Human-Computer Interaction

Franz Madritsch
Inst. f. Computer Graphics, Graz Univ. of Technology
Münzgrabenstr. 11, A-8010 Graz, AUSTRIA
(franz@icg.tu-graz.ac.at)

ABSTRACT

We are investigating the application of CCD-camera based optical-beacon-tracking-systems in 3-d interactive environments. An optical tracking system has been developed which serves as a testbed for tracking algorithms and accuracy investigations. The 3-d interactive environment features tracking of the observer's viewpoints, stereo-scopic visualization and direct 3-d pointing. The focus is set on high accuracy of both, tracking and stereo-scopic visualization. Algorithms which can track optical beacons with sub-pixel resolution at low noise help to reduce hardware expenditure in terms of camera resolution and computing power. Theoretical considerations about resolution are made and practical experience is presented. Furthermore, user studies are performed to test the created interface environment with regard to immersive interfaces and direct interaction in space.

Keywords: optical tracking, interactive visualization in 3-d, input technology, fish tank VR

1. INTRODUCTION

Optical Tracking has already found a variety of applications for specific tracking problems. However, reports about its usage for standard tracking tasks in human computer interfaces are sparse. The main argument against optical methods is poor economy. Viewpoint, head, 3-d cursor tracking, etc. are dominated by commercially available magnetic, acoustic, and mechanical trackers. Deering [Deering, 1992] outlined the demands to create a high fidelity visual interface environment. The requirements concerning tracking accuracy are not met very well with todays magnetic and acoustic tracking devices. Magnetic systems have to deal with high noise and distortion problems especially in the vicinity of CRTs. This is of some importance because of the high retinal resolution achieved by stereo CRT compared to other current VE display technologies [Ware et al., 1993]. Intensive research on optical tracking is done at the University of North Carolina, where several high quality systems have been developed [Ward et al., 1992]. OBT (Optical Beacon Tracking) already has found applications in commercial motion capture systems [AOA, 1994]. Most of them do recording rather than tracking and compute 3-d coordinates off-line. For a comprehensive survey of optical and other tracking systems see also [Meyer et al., 1992].

We shall present a simple optical tracking method which performs exceptionally well in terms of noise, accuracy, and registration. The system is based on optical beacons which are tracked by two fixed color CCD-cameras. Through adapted algorithms, known as centroid calculation in photogrammetry, sub-pixel resolution is achieved [Trinder, 1989, DeMenthon and Fujii, 1994]. These algorithms are computationally inexpensive and lead to a good real-time behavior. Especially for so called "Fish Tank VR" where the user interacts within a limited space in front of the screen, a pair of ceiling mounted cameras can overview the working area perfectly. Positions of interest like shutter-glasses or 3-d cursors can be marked/tracked robustly by e.g. LEDs (light emitting diodes). Through careful determination of the geometric parameters of the cameras and the display screen and their relative alignment to each other, the loop from the eye-points to tracker coordinates and spatial visualization is closed. Accuracies of one part in 3000 are obtained although images with 640 x 480 pixels are used.

2. OPTICAL TRACKING SYSTEM

From the beginning on, we want to avoid specialized hardware and focus our work on the development of fast algorithms. Secondly we want to enable economic software solutions. The system configuration consists of two graphics workstations with live video input, two color-CCD-cameras, and one high resolution computer screen. To avoid high-cost image processing, objects to be tracked are marked with LEDs as optical beacons. Two cameras are fixed at certain positions to overview a defined volume of interest. The whole system rests on the ability to distinguish the beacons from the background and to determine their position within the image. The method of centroid calculation was found to perform this task in a stable and precise manner. The center of a target is defined as the center of gravity of pixels above a certain threshold. Through the right choice of these thresholds in 'RGB' values, red LEDs can be identified robustly in a slightly darkened environment. The centroid calculation is fast and leads to sub-pixel image coordinates of the targets [Madritsch et al., 1996].

Several prototypes were built. Currently our system consists of two CCD-cameras capturing the position of several red LEDs. The position of the beacons in 3-d space is reconstructed out of the two 2-d images provided by the cameras. In principle the position and orientation of the two cameras can be chosen freely; however, to achieve a proper working volume and a precisely measurable geometric configuration, the following setup was found useful.

2.1 Geometric Camera Configuration

The camera coordinate systems are defined so that the y-axes are coincident with the optical axes and the z-axes represent the cameras' view-up vectors. The x-axes complete the right handed Cartesian coordinate systems. The geometry as depicted in Fig. 1 shows the relationship between the two camera coordinate systems, beacons, and the working volume. The motivation for such a set-up is to keep the geometric parameters simple and precisely determinable. To align the xy-planes and the y-axes here, the images of the on-line cameras or even 2-d tracking data can be used for feedback. As a result, the summation of errors is avoided. After that adjustment only three parameters have to be determined by measurement, namely the angle α and the distances from the cameras' first nodal points to their common viewing centers a and b respectively.

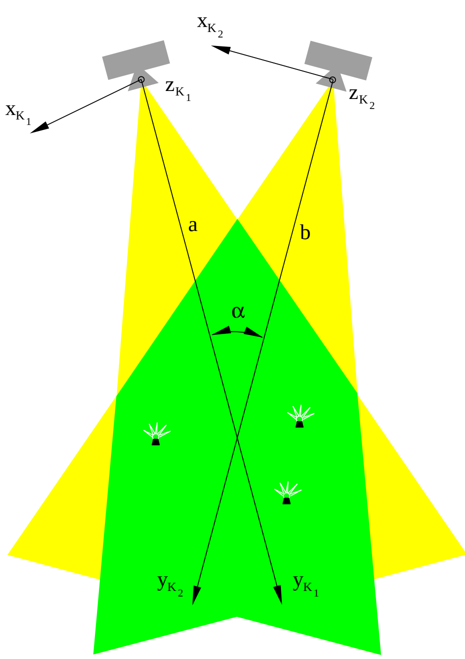

Figure 1: Two cameras with overlapping viewing volume are used to determine the position of several optical beacons

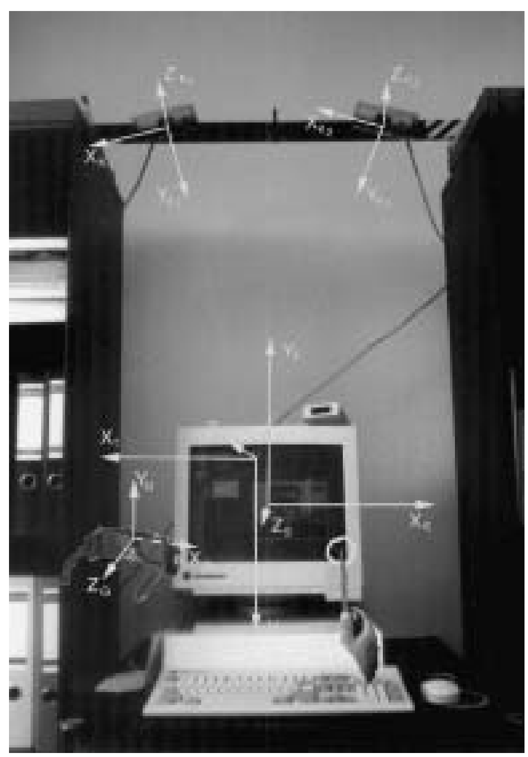

Figure 2: The human-computer interface environment; Two CCD-cameras track the observer's viewpoints and a 3-d pointing device. Visual output via stereo-scopic display.

The position of a beacon in space is determined by inverting the perspective projection equations of each camera [Madritsch and Gervautz, 1996]. Through proper calibration optical distortions of the cameras can be mostly eliminated [Madritsch and Gervautz, 1996, Weng et al., 1992].

2.2 Tracking Algorithm

In general optical tracking of objects is a complex problem the pattern recognition and active vision community has been dealing with for years [Blake and Yuille, 1992, Blake et al., 1993, Broeckl-Fox et al., 1993]. Through easily distinguishable and identifiable beacons marking the objects to be tracked, a method called Optical Beacon Tracking is enabled [Madritsch, 1995, Madritsch and Gervautz, 1996]. Algorithms which are fast and robust can be employed to extract the projections of such beacons out of real-time digital images [Madritsch et al., 1996].

The fish-tank VR system which is described here works with red LEDs as beacons. Two LEDs are placed on top of the stereo-glasses, a third serves as a 3-d pointer. The task of the tracking process is to determine the image coordinates of 3 LEDs within the live-video images. A neighborhood search algorithm is employed in a way similar to the method described in [Madritsch et al., 1996]. A quadratic region of fixed side length is searched. As an estimate of the beacon position the center of the search-region is located at the last known beacon position. So far no prediction is used. In the basic step the whole image is searched till all three beacons are found. In subsequent steps corresponding to subsequent frames only the neighborhood search-region is considered first. If the neighborhood search fails for one or more beacons, the remaining image is searched line by line causing a temporal distortion in the flow of the tracking data due to the increased search time. Fig. 3 shows an image of three beacons as seen by one camera.

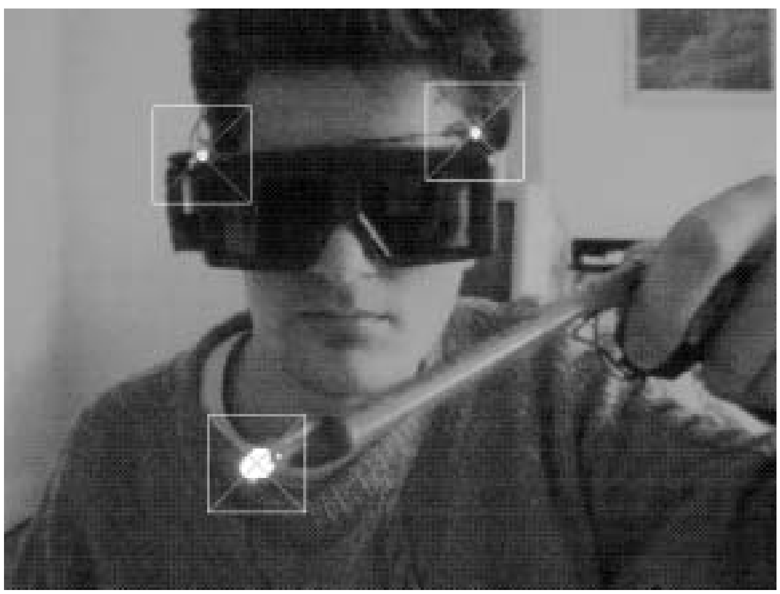

Figure 3: Optical beacons from the viewpoint of the camera. Extracted centroids deliver sub-pixel image coordinates.

The size of the neighborhood search-region influences both the maximum trackable beacon speed and the search time and consequently the update rate. Reflections about the optimal search-region size and alternative algorithms have been made in [Madritsch et al., 1996].

In the fish-tank application there are two kinds of beacons with different temporal behavior; the two LEDs connected to the shutter-glasses worn by the user, and one LED on top of the 3-d pointer. From data recorded during several tests and our practical experience during the evaluation of our prototype we found that the neighborhood assumption is usually correct. However, rapid user motions may cause a temporal loss of the beacons. This case occurs when the 3-d pointer is moved abruptly rather than when the user's head is moved. The spatial impression of the scene is usually maintained steadily.

2.3 Software system configuration

Within our environment a usual optical tracking task consists of three processes: Two identical image processing tasks extract the beacons in a neighborhood search and perform the centroid calculation (see Tracking Algorithm). One further process combines the image coordinate pairs to obtain Cartesian 3-d tracker coordinates and delivers position and orientation data to an update and rendering process. In our implementation the image processing runs on two separate workstations. To avoid the transmission of images over the network image processing is done locally on the workstation where the frame-grabbing occurs. All interprocess-communication is currently done by Berkeley datagram sockets. Such data links are fast but not completely safe. Receiving real-time position data without delay is, however, usually as important as its correctness.

2.4 Accuracy

In OBT-systems, as in most other optical systems, the absolute accuracy is determined by the system's "resolution" and the working volume because accuracy can always be scaled by the focal length of the cameras. The term "resolution" must be treated with care in OBT because the sizes of the tracking steps vary. We would like to define resolution as the reciprocal value of the uncertainty multiplied by the range. (and not by the number of distinguishable steps) There can be big discrepancies between these definitions in OBT. The resolution of a OBT-system is given by the resolution of the camera multiplied by a sub-pixel factor. Through centroid calculation the image coordinates of the beacons can be determined with sub-pixel accuracy. The value of the sub-pixel factor is mainly dependent on the relative beacon size within the digital image. In the fish-tank application we worked with a sub-pixel factor of about 4 to 6. Fig. 4 shows how a sub-pixel region is resolved by unweighted discrete centroid calculation for a circular target of 4 pixels in diameter. This pattern is the result of a simulation. A detailed investigation of the discrete centroid calculation together with results from physical measurements can be found in [Madritsch et al., 1996].

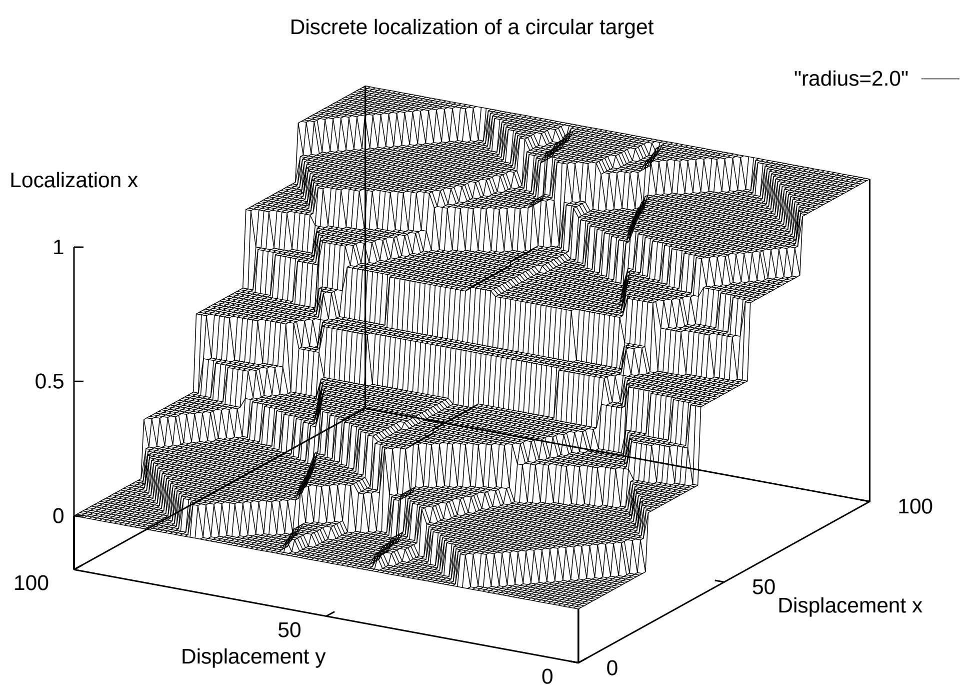

Figure 4: Simulation of target localization by discrete centroid calculation for a relative target diameter of 4 pixels (100 x 100 steps; "Localization x" in grid units)

Using CCD-cameras with a resolution of 640 x 480 pixels a two dimensional tracker resolution of 3200 x 2400 is reached. The resolution in the third coordinate direction is similar but differently scaled depending on the geometric camera arrangement. Our configuration as depicted in Fig. 1 with the parameters $a = 85cm; \quad b = 85cm; \quad \alpha = 36°$ led to tracker accuracy values of 0.17 x 0.18 x 0.6 mm RMS. The absolute registration accuracy between tracker and spatial visualization is ± 1 mm as evaluated by visual registration tests.

3. SPATIAL COMPUTER INTERFACE

Interfaces with 3-dimensional input and output capabilities are becoming more and more common. Perhaps the biggest shift in public consciousness happened when J.D. Foley published their work at NASA [Foley, 1987]. Since the extensive press and media coverage in 1990-1991 the term "Virtual Reality" has become common currency. The historical evolution and state of the art is summarized in [Krueger, 1990, Rheingold, 1991, Ellis, 1991, Kalawsky, 1993]. Stereoscopic visualization and 3-d pointing are small fields within this huge area but nevertheless key technologies in 3-d interfaces. The fish-tank concept used here for spatial visualization is not immersive but offers high retinal resolution to the user and meets the criteria for high definition VR as outlined by [Deering, 1992] quite well. The tracking task done by OBT also fits in, although there is room for improvements in temporal performance as well as in the performance of the graphics workstation. In the following sections a spatial visualization system will be outlined which generates a limited volume where virtual objects can be placed and viewed by one observer wearing tracked stereo-glasses. Furthermore a tracked 3-d pointer enables the observer not only to view the scene passively but also to interact with it in a certain way. Let us first take a look at perceptual issues of spatial visualization.

3.1 Depth Perception and Stereoscopic Viewing

Mostly based on our visual sense the human perceptual system builds a "model in mind" of the 3-dimensional world around us. There are various visual depth cues which allow us to estimate distances and spatial relations with different certainty and reliability. Motion parallax, stereopsis, linear perspective, size constancy, and accommodation are some of the most important visual depth cues [Storey and Craine, 1985]. Most stereoscopic viewing techniques cannot reproduce all of them. In the case of stereoscopic color displays accommodation is always bound to the distance between the eyes of the observer and the screen in order to produce sharp images on the retinae. Stereopsis is achieved by presenting each eye with an image in a perspectively correct way. By using perspectively correct images the "size constancy" depth cue is also reproduced. To calculate correct stereo pairs the position of the observer relatively to the display device must be known. This position must either be assumed to be constant or tracked [Meyer et al., 1992]. In the second case, the rendering routine can use real-time information about the observer's viewpoints to produce current stereo-pairs on-line. This method further enables the depth cue of observer induced motion parallax. Other depth cues such as lighting, shading, atmospheric effects, hidden surface removal, etc. are matters of scene description and graphics performance. In conclusion it can be said that a color stereo screen combined with tracked goggles can principally present the observer with an impression which has almost all visual properties a real object or scene would have. One problem that cannot be solved in this way is the accommodation-convergence breakdown. The eyes converge at the point in space where the virtual object appears while still focused to the screen. This effect troubles the learned coordination between the muscles moving the eyes and the ciliar muscles focusing the lenses. The result of this discrepancy is inconvenience of the viewer when looking at stereo pairs with big vertical parallaxes. Therefore stereo-visualizations should inhabit only a limited volume in front of and behind the screen. According to [Lipton, 1993] both positive and negative parallax should not exceed 1,5°. Other restrictions are finite resolution, imperfect color and brightness reproduction and limited scene detail, as well as temporal performance restrictions. Mathematics of stereo image computations are given in [Deering, 1992, Hodges and McAllister, 1993].

The importance of different depth cues was studied by [Ware et al., 1993]. In their investigation, several subjects were asked to trace 3-dimensional tree structures from certain branches to the corresponding root. The viewing conditions were: monocular, stereo only, head-coupled only, stereo and head-coupled. Although the test persons' subjective opinion preferred head-coupled only over head-coupled and stereo, the timing and error data of the experiments showed that performance increased from stereo to head-coupled only and was best with head-coupled stereo.

Especially in the grasp-range of a user the binocular disparity and vergence (stereopsis) provide an almost absolute cue for depth estimation. Short range depth estimation is used in every-day life for every grasping action and therefore this ability is highly trained. The advantage of stereopsis compared to motion parallax is the instant spatial impression provided without the need of an exploring motion.

3.2 Position of the user's eye-points

For precise registration of virtual and physical objects the position of the first nodal points of the eyes has to be tracked in an ideal case [Deering, 1992]. On the other hand in [Madritsch, 1995] it has been shown that very little information, namely only the direction vector from the display to the user is sufficient to produce eye-point corrected stereoscopic image pairs. In practice it proves to be difficult and technically expensive to track the eye-points (an additional tracker for gaze direction would be necessary [Starks, 1991]). The knowledge of the gaze direction only adds little to the precision of the position of the eye-points since the center of rotation of the eye and its first nodal point lie closely together and the user looks straight at the screen rather than with her/his head turned away from it. For our fish tank application we use the eye-point model described in [Madritsch et al., 1996]. Two LEDs connected to the stereo-goggles are tracked 3-dimensionally. The geometric relationship between the markers and the eye-points is defined in a head aligned coordinate system, see Fig. 5. The vectors from the LEDs to the eye-points are determined from an average head size. Due to the fact that the transformation from tracker to head coordinates cannot be stated completely by two tracked points, the user's head is assumed to be in an upright position. Nodding motions would induce errors depending on the separation between the beacons and the eyes. In practice this error can be neglected.

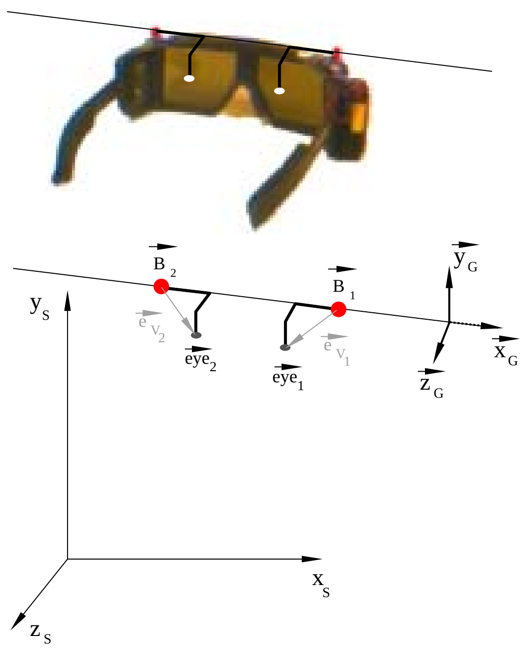

Figure 5: Geometric model of the relation between beacons and eyes

$$\vec{eye_i} = \vec{B_i} + \vec{e_{v_i}} \qquad i = 1,2 \tag{1}$$

with

$$\vec{e_{v_i}} = M\vec{e_{v_i}} \tag{2}$$

The transformation matrix M is obtained by expressing the goggle coordinate system's basis vectors $\vec{x_G}$, $\vec{y_G}$, and $\vec{z_G}$ in system coordinates

$$R = (\vec{x_G} \quad \vec{y_G} \quad \vec{z_G}) \tag{3}$$

where

$$\vec{x_G} = \frac{\vec{B_1} - \vec{B_2}}{\left|\vec{B_1} - \vec{B_2}\right|}, \tag{4}$$

$$\vec{y_G} = \vec{z_G} \times \vec{x_G}, \tag{5}$$

$$\vec{z_G} = \left(-\frac{x_{G_z}}{\sqrt{x_{G_x}^2 + x_{G_z}^2}} \quad 0 \quad \frac{x_{G_x}}{\sqrt{x_{G_x}^2 + x_{G_z}^2}} \right)^T. \tag{6}$$

Note that $\vec{y_G}$ is kept aligned with the $y_S - axis$ due to the missing rotational information around the $x_S - axis$.

3.3 Registration of virtual and physical objects

To calculate current perspectively correct image pairs the position of the eye-points and visualization coordinates must be expressed in the same coordinate system. We chose a screen oriented coordinate system as the common basis for all coordinates. Such coordinates will be referred to as system coordinates.

A priory the transformation between tracker coordinates and visualization coordinates is unknown. In our case where the tracker supports 3-d pointing it suggests itself to establish a connection between these coordinate systems by selecting certain points on the display as anchors for the transformation. We use the 3 corners P_1, P_2, P_3 of the usable screen area as anchor points, see Fig. 6. Since these anchor points also correspond to the physical viewport limits they can be used to scale the visualization to physical units, provided that the output data of the tracker is calibrated properly.

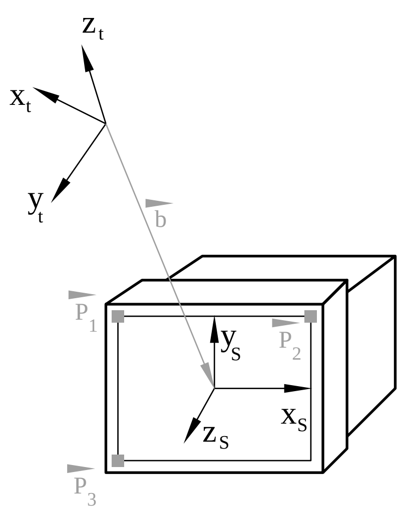

Figure 6: Geometric model of the relation between tracker coordinates and visualization (system) coordinates. For simplicity all coordinates (tracking related and visualization related) are transformed into system coordinates.

Tracker coordinates transform into system coordinates by

$$\vec{P_S} = S^T \vec{P_t} - S^T \vec{b} \tag{7}$$

with

$$S = (\vec{x_s} \quad \vec{y_s} \vec{z_s}), \qquad \vec{b} = \frac{\left(\vec{P_2} + \vec{P_3}\right)}{2}, \tag{8}$$

and

$$\vec{x_S} = \frac{\vec{P_2} - \vec{P_1}}{\left|\vec{P_2} - \vec{P_1}\right|}, \qquad \vec{y_S} = \frac{\vec{P_1} - \vec{P_3}}{\left|\vec{P_1} - \vec{P_3}\right|}, \qquad \vec{z_S} = \vec{x_S} \times \vec{y_S}. \tag{9}$$

This semiautomatic "self-adjustment" is done by the user by pointing at the corners P_1, P_2, P_3. It is only necessary after changes in the arrangement between display device and tracker have occurred.

3.4 Interacting in 3D

Venolia [Venolia, 1993] showed how to build up an interface for direct interaction in 3-d using only a minimum set of interface elements. The system was based on the "roller mouse" (a mouse enhanced with additional wheels to control the third dimension as well). Various 3-d input devices and techniques such as a Bat (flying mouse) [Ware and Jessome, 1988], a data glove [Sturman and Zeltzer, 1994] or a SPACEBALL[1] [Zhai, 1993], just to mention a few, have been investigated. This section will mainly deal with 3-d pointing and some of our experience with this kind of interface.

The system described above establishes a "virtual volume" where computer generated 3-d objects can be placed at will. This volume is a limited space in front of and behind the display screen. The condition for stereoscopic visibility is that from the view of each eye the entire virtual object can be "projected" onto this screen. A three dimensionally tracked stylus allows the user to interact directly within this volume.

3.4.1 3D Pointing We use an LED mounted on a pencil together with the supplying battery as a 3-d pointing device. The position of the LED is tracked and available in system coordinates. A virtual cursor can be placed at the identical position in space. (In this way the registration of physical and virtual coordinates can be checked easily.) In an application where a racket of a squash game was controlled by the 3-d pointer directly the end of the virtual handle was aligned with the pointer. It turned out to be disturbing when the virtual representation of the pointer was obscured by its physical counterpart. If proper registration is given no representation of the physical pointer is needed. But a virtual representation gives confidence to the user that she/he really points to the place in the virtual scene where it appears. This feedback seems to be helpful especially to novice users who are usually surprised that they can really "touch" a virtual object. So we found it useful to simply extend the physical pointer by a virtual top. By further extending the physical pointer the space behind the visualization screen can also be used for interaction, thereby increasing the working volume. It is not completely clear whether a direct one to one correspondence between physical and virtual pointer improves the user's pointing performance. The user might, for example, also control the virtual cursor remotely in a way similar to the situation of controlling a 2-d cursor by a 2-d "mouse". However, direct interaction seems to be more intuitive.

3.4.2 Interactions We implemented several applications utilizing the fish tank VR environment to perform simple interaction experiments. The 3-d squash game mentioned above was one of them. This example where the user can only hit a bouncing ball by a racket controlled by the 3-d pointer showed that a naive user can understand and adapt to such a scenario without any instruction.

As a student project a simple 3-d editor was designed. Models consisting of quaders and ellipsoids were created, stored, and reloaded. Interactions such as moving, scaling, and rotating were supported. Both-handed input was used. One hand controlled the stylus, one hand the "mouse" to select different modes. Manipulations were performed by selecting an object with the pointer and choosing either the move, the scale, or the rotation mode. Moving was done by connecting the center of mass of the object with the pointer and releasing the object at the desired place. Scaling was done by controlling a corner of the object's bounding box while its center stayed fixed. To control the rotation over all 3 axes simultaneously by the pointer is not so straight forward since the "handle on the surface" concept only allowed rotations around 2 axes. In a second approach the mouse which was used as a selection device was replaced by a SPACE MOUSE[2] (6 degree of freedom input device similar to a SPACEBALL™). It improved user performance to select, move, and scale objects directly with the pointer and rotate objects and the scene directly with the SPACE MOUSE ™. We learned that such an interface allows very intuitive handling and manipulation of the objects, requiring little instruction for novice users.

In our opinion both-handed input should be used in such an environment and the user's hands should stay at the input devices. There should be no need of looking away from the center of visual attention while interacting in 3-d. We think that visual discontinuities are even more disturbing in 3-d interfaces than they are in 2-d interfaces. Because of the high number of degrees of freedom it is difficult to position and align objects in space exactly. It is more difficult to align objects in depth [Zhai and Milgram, 1994]. Sometimes it is therefore useful to rotate the whole scene. There is a great need for methods to support the user.

The user tests were performed by observing and interviewing about 50 visitors the system was demonstrated to.

---

[1] SPACEBALL is a trademark of Spaceball Technologies, Inc.
[2] SPACE MOUSE is a trademark of Deutsche Forschungsanstalt für Luft- und Raumfahrt

4. PERFORMANCE FIGURES

resolution: 3200 x 2400 x 2000 volume elements
accuracy: 0.17 x 0.18 x 0.6 mm RMS
absolute registration: $\pm$ 1 mm
update rate: 30 $\frac{1}{s}$
lag: 40-60 ms
scene complexity: several hundred polygons flat shaded and wire-frame

The absolute registration deviations were determined by visually checking the alignment of a physical 3-d pointer and a virtual cursor within the closed tracking visualization loop of the fish-tank environment.

All results are based on a hardware configuration consisting of two workstations with MIPS™ R4600 100 MHz processors with live video input, two color CCD-cameras (640 x 480 RGB-pixels resampled out of 512 x 492 pixels), one SPACE MOUSE™, and one stereo monitor with LCD shutter-glasses.

5. CONCLUSIONS

The OBT concept offers high resolution and accuracy at acceptable temporal performance. The line of sight restriction of the optical system represents no problem for a fish tank VR configuration. It supports the tracking of the observer's viewpoint and a 3-d stylus robustly. Probably the main advantage of OBT compared to magnetic methods is that there is practically no noise in the tracking data (i.e. the noise is within the error bounds) as long as the neighborhood search works smoothly. Optical beacons are hardly intrusive. No cabling is needed for the goggles and the stylus.

A fish tank VR environment can serve as an intuitive three dimensional interface. We think that today it might be appropriate for naive users rather than for experts because simple tasks can be done very easily. But even in advanced VE systems it might turn out that the more complicated and complex the interface becomes the more of the advantages of a direct interface are lost.

6. ACKNOWLEDGMENTS

Many thanks to Stefan Brantner, Peter Duracher, Harald Ganster, Michael Gervautz, Michael Gruber, Konrad Karner, Franz Leberl, Klaus Wiltschi, and Andrea Zunegg for their support and contribution.

This research project was funded by the ADW under project number 6070/16-II/5/95.

7. REFERENCES

[AOA, 1994] (1994). *Capture Motion, Release Creativity (Trade Literature).* United Technologies Adaptive Optics Associates, 54 CambridgePark Drive, Cambridge, MA 02140-2308.

[Blake et al., 1993] Blake, A., Curwen, R., and Zisserman, A. (1993). A framework for spatio-temporal control in the tracking of visual contours. *Int. Journal of Computer Vision*, 11(2):127–145.

[Blake and Yuille, 1992] Blake, A. and Yuille, A. (1992). *Active Vision.* Massachusetts Institute of Technology Press.

[Broeckl-Fox et al., 1993] Broeckl-Fox, U., Hartensteiner, K., Kettner, L., Nolting, O., G.Selinger, Weiss, B., and Schmitt, A. (1993). Using three dimensional hand-gesture recognition as a new 3d input device. Technical report, Institut fuer Betriebs- und Dialogsysteme; Universitaet Karlsruhe, Am Fasangarten 5; Postfach 6980; Karlsruhe.

[Deering, 1992] Deering, M. (1992). High Resolution Virtual Reality. *Computer Graphics*, 26(2):195–202.

[DeMenthon and Fujii, 1994] DeMenthon, D. F. and Fujii, Y. (1994). Three dimensional pointing device monitored by computer vision. U.S. Patent 5,297,061, (Application 08/063489, May 1993).

[Ellis, 1991] Ellis, S. R., editor (1991). *Pictoral communication in virtual and real environments.* Taylor & Francis.

[Foley, 1987] Foley, J. D. (1987). Interfaces for advanced computing. *Scientific American*, 257(4).

[Hodges and McAllister, 1993] Hodges, L. F. and McAllister, D. F. (1993). *Stereo Computer Graphics and Other True 3D Technologies*, chapter 5, pages 71–89. Princeton University Press, 41 William Street, Princeton, New Jersey 08540.

[Kalawsky, 1993] Kalawsky, R. S. (1993). *The Science of Virtual Reality and Virtual Environments.* Addison-Wesley.

[Krueger, 1990] Krueger, M. W. (1990). *Artificial Reality II.* Addison-Wesley.

[Lipton, 1993] Lipton, J. (1993). *Composition of Electrostereoscopic Displays*, chapter 4, pages 11–25. Princeton University Press, 41 William Street, Princeton, New Jersey 08540.

[Madritsch, 1995] Madritsch, F. J. (1995). Correct spatial visualization using optical tracking. *Virtual Reality: Research, Development & Applications*, 1(2):122–126.

[Madritsch and Gervautz, 1996] Madritsch, F. J. and Gervautz, M. (1996). Ccd-camera based optical beacon tracking for virtual and augmented reality (in print). *COMPUTER GRAPHICS forum*, 15(3).

[Madritsch et al., 1996] Madritsch, F. J., Leberl, F. W., and Gervautz, M. (1996). Camera based beacon tracking: Accuracy and applications (in print). *ACM Transactions on Human-Computer Interaction*, 3(3).

[Meyer et al., 1992] Meyer, K., Applewhite, H. L., and Biocca, F. A. (1992). A Survey of Position Trackers. *PRESENCE*, 1(2):173–200.

[Rheingold, 1991] Rheingold, H. (1991). *Virtual Reality.* Summit Books / Simon & Schuster, New York.

[Starks, 1991] Starks, M. (1991). Stereoscopic video and the quest for virtual reality: an annotated bibliography of selected topics. *SPIE Proceedings, Stereoscopic Displays and Applications II*, 1457:327–342.

[Storey and Craine, 1985] Storey, N. and Craine, J. F. (1985). Interactive Stereoscopic Computer Graphics Display Systems. In Shackel, B., editor, *Human-Computer Interaction - INTERACT '84*, pages 163–168. Elsevier Science Publishers B.V., Amsterdam.

[Sturman and Zeltzer, 1994] Sturman, D. J. and Zeltzer, D. (1994). A survey of glove-based input. *IEEE Computer Graphics & Applications*, 14(1):30–39.

[Trinder, 1989] Trinder, J. C. (1989). Precision of Digital Target Location. *Photogrammetric Engineering and Remote Sensing*, 55(6):883–886.

[Venolia, 1993] Venolia, D. (1993). Facile 3D Direct Manipulation. In *INTERCHI '93*, pages 31–36.

[Ward et al., 1992] Ward, M., Azuma, R., Bennett, R., Gottschalk, S., and Fuchs, H. (1992). A demonstrated optical tracker with scalable work area for head-mounted display systems. In *SYMPOSIUM ON INTERACTIVE 3D GRAPHICS*, Special issue of Computer Graphics, pages 43–52. ACM SIGGRAPH.

[Ware et al., 1993] Ware, C., Arthur, K., and Booth, K. S. (1993). Fish Tank Virtual Reality. In *INTERCHI '93*, pages 37 – 42.

[Ware and Jessome, 1988] Ware, C. and Jessome, D. R. (1988). Using the Bat: a six-dimensional mouse for object placement. *IEEE Computer Graphics & Applications*, 8(6):65–70.

[Weng et al., 1992] Weng, J., Cohen, P., and Herniou, M. (1992). Camera calibration with distortion models and accuracy evaluation. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 14(10):965–980.

[Zhai, 1993] Zhai, S. (1993). Investigation of feel for 6dof inputs: isometric and elastic rate control for manipulation in 3d environments. In *PROCEEDINGS of the HUMAN FACTORS AND ERGONOMICS SOCIETY 37th ANNUAL MEETING.*

[Zhai and Milgram, 1994] Zhai, S. and Milgram, P. (1994). Asymmetrical spatial accuracy in 3d tracking. In *PROCEEDINGS of the HUMAN FACTORS AND ERGONOMICS SOCIETY 38th ANNUAL MEETING.*