7. International Immersive Projection Technologies Workshop
9. Eurographics Workshop on Virtual Environments (2003)
J. Deisinger, A. Kunz (Editors)

An Affordable Optical Head Tracking System for Desktop VR/AR Systems

Jurriaan D. Mulder, Jack Jansen, and Arjen van Rhijn

Center for Mathematics and Computer Science, CWI
Amsterdam, the Netherlands
{mullie,jack,arjenvr}@cwi.nl

Abstract
We present an affordable optical head tracking system for desktop-like VR/AR environments. The generic and specific head tracking requirements for these type of environments are defined, as well as the relaxations such environments put on head tracking systems. The presented head tracker is based on two low-cost, commodity FireWire cameras that track a simple 3D dot pattern. It is shown that the tracker provides high accuracy, an update rate of 30 updates per second, a low computational load, and a moderate delay of 66 ms. It is competitive to commercially available, moderate-cost head tracking systems yet for substantially lower costs.

Keywords: Optical Head Tracking, Desktop Virtual and Augmented Reality.

Categories and Subject Descriptors (according to ACM CCS): I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Stereo, Tracking; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual reality; H.5.2 [Information Interfaces and Representation]: User Interfaces—Evaluation/methodology, Input devices and strategies;

1. Introduction

Over the past number of years, there is an increased interest toward more affordable environments for virtual and augmented reality (VR/AR). Many types of 3D applications do not require CAVE-like, fully immersive environments. Often, *fish tank, desktop,* or *dexterous* types of environments are sufficient [8, 19, 26]. For a major part, these types of environments can be constructed out of commodity, off the shelve hardware, lowering the costs significantly. At the Center for Mathematics and Computer Science, an affordable desktop environment for near-field VR/AR is being developed; the Personal Space Station (PSS) [16]. The emphasis of the PSS is on *direct 3D interaction, ergonomics,* and *low costs*. We believe that by creating an affordable environment that is comfortable, easy, and intuitive to use the acceptance threshold of VR/AR can be lowered significantly.

Head tracking is a crucial factor in VR/AR. In fact, user dynamic control of the viewpoint can be considered a requirement for any (3D) application to be a *virtual reality experience* [6]. Furthermore, it has been shown that head tracking can increase 3D task performance significantly [2, 26]. Many types of head tracking systems have been developed over the years [3, 27], based on electro-magnetic, optical, inertial, acoustical, and mechanical sensing techniques. However, the number of affordable (say, less than 1 k€) yet adequate head tracking systems is limited.

In this paper, we present an affordable optical head tracking system that uses two low-cost, standard FireWire cameras. We will show that the system can compete with commercially available systems such as the Logitech acoustical head tracker, yet for significantly lower costs.

In the next section, we will elaborate on the background and motivation for developing the system, and formulate the requirements and variables involved in head tracking for desktop VR/AR systems. In Section 3 we describe the hardware and software configuration and implementation of our tracking system. In Section 4 the performance of the tracking system is evaluated and compared to that of the Logitech acoustical tracker.

2. Background

2.1. The Personal Space Station

The PSS is designed with three major goals: low costs, ergonomics, and direct 3D interaction. A schematic diagram of the PSS is depicted in Figure 1. The PSS uses a mirror-based display, as has also been successfully applied in other systems [19, 23, 28]. The user is seated in front of the mirror which reflects the stereoscopic images of the virtual world as displayed by the monitor. Different types of mirrors can be used: a fully reflective mirror for pure VR applications, or a half-silvered mirror to mix real and virtual objects and create an AR environment.

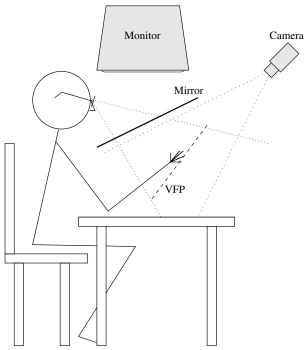

Figure 1: *Schematic side view of the Personal Space Station.*

The main advantage of this type of display set-up is that the visual space and the interaction space are allowed to coincide: The user can reach *under* the mirror *into* the virtual world without obscuring the image or colliding with the monitor. As a result, proprioception and hand-eye coordination are exploited, and interaction becomes more direct. A prototype PSS is depicted in Figure 2.

A unique aspect of the PSS is that it is designed to be a modular unit that can be used to construct a collaborative VR/AR environment with a shared workspace [15]. By cascading multiple PSS's together such that their visual and interaction spaces overlap, a *Joint Space Station* is created. Each user has a correct stereoscopic view and can perform direct 3D interaction in the shared physical and virtual workspace.

Figure 2: *A prototype Personal Space Station.*

2.2. Head Tracking

2.2.1. Generic Requirements

Any head tracking system to be used in a VR environment has to be accurate, of low latency, provide sufficiently high update rates, and be robust. In an AR environment these aspects are even more critical; any head tracking deficiencies will result in unrealistically behaving objects (e.g. 'swimming' and deforming), which is especially noticeable when perceived simultaneously with real, physical objects in the scene that do behave normally.

Accuracy: Accuracy is of utmost importance. It has been stated that in the ideal case positional tracking errors should not exceed 1 mm and orientational errors should be below 0.1 degree [22, 27]. Head tracking errors have different effects depending on the type of display used. With head-mounted displays, positional and rotational errors will cause objects to be rendered at faulty positions. Jitter in orientation can result in a vigorously moving scene. When stationary displays are used positional and rotational errors can cause objects to deform. However, rotational errors and jitter are less of a problem with stationary displays [7, 27].

Low-latency: The overall latency (or lag) in a head tracked display system can be defined as the period in time from the actual sampling of the head position to the actual displaying of the scene rendered according to that head position. The amount contributed to the overall delay by the head tracking system is the duration from the sampling of the actual head position to the point in time where this head position has been calculated and made available to the display system. Although lag in stationary displays is less problematic than in head-mounted displays, overall lag should be kept to a minimum. Preferably, perceived lags should be no more than 50-100 ms [8].

Update Rate: The update rate of the head tracker should meet (or exceed) the display frame rate. Most sequential stereo display systems run at about 60 frames per second. A display frame rate of 10 Hz is often considered to be the minimum to achieve smooth animation. However, a display frame rate of 10 Hz can introduce an effective lag of 175 ms which can have a significant effect on 3D task performance [25]. If an accurate head pose is to be determined at all times, then according to Shannon's theorem the head pose sampling frequency should at least be twice the highest frequency of motion of the head.

Robustness: Obviously, the tracker should be as robust as possible. However, during an interactive session, it will be less problematic when an occasional head position is missing and the head position from the last update is used (perceived as an increased lag) than when utterly wrong head positions are reported (resulting in flashing or jumping of the scene).

2.2.2. Specific Requirements and Relaxations

Since we want to use the head tracker in the PSS, a number of additional requirements have to be met, partially following from the design philosophy of the PSS. However, the envisioned use also relaxes a number of constraints that might be more strict for other head tracking applications. We describe the most important requirements and relaxations. Many of these also apply to other desktop-type of VR environments.

Limited Range: Head movements in desktop VR are limited. In the PSS, a tracked volume of about 0.5 m wide, 0.3 m high, and 0.4 m deep suffices. Head rotations are limited as well. Neither yaw, pitch, nor roll rotations are likely to exceed 30 degrees each way.

Single 6D Sensor: The tracker will be dedicated to tracking head position and orientation; it will not be used for the tracking of other interaction devices. Furthermore, only a single user will be tracked. Therefore, the tracker only has to be capable of tracking one single '6D sensor'.

Interference Free: The PSS is to be used under office working conditions. Furthermore, each PSS is to function as a stand-alone system, but they should also be cascadeable to create a Joint Space Station. Therefore, the head tracker should not be sensitive to noise or specific materials in the near surrounding, nor should it interfere with the tracking systems of other PSS's. On the positive side, a free line of sight from within a PSS to the user's head is available and occlusion problems are not likely to occur.

Low Cost: The PSS is designed as a low-cost VR system. Therefore, the head tracker should be low-cost as well, preferably not more than a couple of hundred euros.

Low Computational Load: Partially because of cost reduction, the PSS is aimed to be driven by a single standard PC. All necessary software (application, graphics, interaction and head tracking) will reside on this PC. Therefore, the computational load put on the system by any head tracking software should be as low as possible.

Non-intrusive: The PSS is designed to be ergonomic and easy to use. Therefore, any device to be used should be as little intrusive or hindering to the user as possible.

Compact: The PSS is a compact design. In order to keep the PSS physically manageable and to be able to cascade multiple PSS's together without physical interference, there shouldn't be any devices extruding from the PSS chassis.

Simple: Using and configuring the PSS, and thus the head tracker, should be easy.

2.2.3. Logitech Tracker

Initially, the prototype PSS's were equipped with a Logitech acoustical head tracking system. This commercial product is readily available for about 2 k€. The Logitech device however, does not fulfill all of the above requirements:

- The receiver mounted on top of the shutter glasses is too heavy making the glasses uncomfortable to use.
- The device is not wireless which hinders the user.
- The active range of the tracker is limited to the overlap of the 100 degree spherical cones emerging from the three speakers. As a result, the emitter must be located too far away from the user to allow a truly compact construction of the PSS.
- The display monitor, mirror, and the PSS chassis can cause reflections of the ultrasonic signals resulting in faulty measurements.
- The tracker is sensitive to sounds in the near environment (jingling keys for instance).
- When cascading multiple PSS's together to construct a Joint Space Station, the Logitech tracking devices of the different Space Stations may interfere.

2.3. Optical Tracking

Potentially, optical tracking is very well suited for head tracking in the PSS or any other desktop-like VR system. Several optical tracking systems are available [21], ranging from high-end ready to use commercial products to public domain software systems. Optical tracking systems can be classified in two categories: *marker* based and *image* (or *feature*) based.

2.3.1. Marker Based

These type of systems make use of active or passive markers that are attached to the object(s) to be tracked. The major advantage of this approach is that by choosing these markers appropriately they can be found relatively easily in the camera images. This is particularly the case in systems that use infrared light for the active or passive (reflective) markers. In this case, image processing comes down to simple blob detection in a greyscale image. This approach is also used in the PSS for the tracking of interaction devices [13].

Several commercial marker based optical tracking systems exist that provide sufficient performance to be used for head tracking in VR/AR. Examples include the OPTOTRACK and POLARIS tracking systems from NDI. Most of these commercial products however are far too expensive to be considered for use in the PSS. At the lower end, the commercial optical systems do not meet our requirements. An example of such a low end tracking system is the DynaSight Sensor system available from Origin Instruments for a little over 2 k€. This system uses passive or active IR markers. However, the system is only capable of providing the 3D position of a single wireless marker. To obtain a full 6D sensor, multiple active markers will have to be used that have to be synchronized with the base unit via a cable. Other low-end systems such as the trackIR system from NaturalPoint or the Tracker One and Tracker 2000 systems from Madentec use a single camera and can track only a single passive or active IR marker. These systems are not suited for real VR/AR applications as they cannot provide 6D sensor information, sufficient resolution, nor adequate update rates. These systems are mostly used for head-induced 2D cursor control.

Several IR marker based tracking systems have been constructed at various research institutes [9, 13, 22]. Most of these systems however, require custom-build hardware and high-end cameras and frame grabbers which are not readily available or too expensive to meet our requirements.

2.3.2. Image Based

In image based systems more complex image processing is needed to extract certain features (such as corners or edges) of the objects to be tracked out of the obtained images. Much research has been done in this area, where most research is aimed at face [11, 24], eye [1], or head [5, 12, 20, 30] tracking using a single camera, or dual cameras [10, 14]. However, none of these approaches as of yet delivers the required performance with respect to update rate, robustness, and/or accuracy. Especially when a single, possibly low-end camera is used, accuracy is limited, in particular for distance measurements. Using dual cameras can increase the accuracy, but doubles the processing time of the feature detection resulting in high computational loads, low update rates, and high lags. Again, most of these systems are used for 2D cursor control or target selection.

3. System Overview

3.1. Hardware

Our tracking system uses two FireWire iBOT cameras [18], which are readily available for a little under 100 €. The cameras in conjunction with the FireWire interface allow us to grab uncompressed, progressive-scan (non-interlaced) 640x480 sized greyscale images at 30 frames per second from both cameras through DMA. The cameras are manual focus and several parameters such as brightness, gain, aperture and shutter can be adjusted through software to create sharp images with minimal motion blur.

The cameras are mounted to the PSS chassis in front of the user. A simple marker pattern has been attached to the shutter glasses worn by the user, as depicted in Figure 3. The marker pattern consists of 3 black circular dots on a white background arranged in a triangular pattern. Three dots is the minimum to obtain full position and orientation data of the marker pattern. More dots could be added to increase stability towards imprecise measurements. The (known) inter-dot distances can be used to search for the marker pattern in a cloud of candidate 3D marker positions.

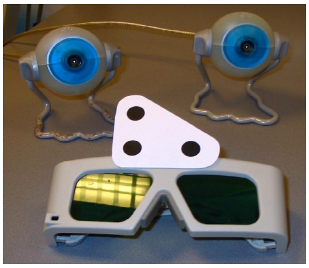

Figure 3: *The tracking system hardware: Two iBOT FireWire cameras and the dot pattern (mounted onto the shutter glasses).*

3.2. Algorithm

The basic principles in the tracking algorithm are:

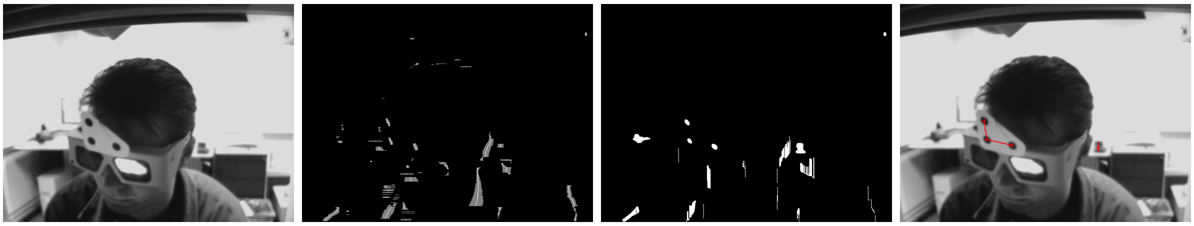

Figure 4: *Algorithmic steps. From left to right: the input image, the light-dark-light regions found by the edge detector, the result of the flood fill, and the ellipses found in the original image (indicated with red crosses).*

1. Identifying candidate dot markers in the left and right images.
2. Matching left and right candidate dots and reconstructing their 3D position.
3. Searching among 3D candidate positions for the correct dot pattern.

These three basic principles are interleaved. That is, they function in an on-demand basis. A pool of left and right candidate dots is maintained, along with the 3D positions of matching pairs among these candidate dots. As long as the desired 3D pattern is not found among the 3D positions in the pool, new candidate dots are searched for in the left and right images. These candidate dots are added to the pool and the 3D positions of any new matching pairs are computed. It is checked whether the marker pattern can be found with the newly added candidate 3D positions. This process continues until either the 3D pattern is found or no more candidate dots are available in the left and right images.

3.2.1. Searching Candidate Dots

Three basic steps are performed to find candidate dots in the images: *edge detection, flood fill,* and an *ellipsoid check*. These steps are illustrated in Figure 4. When searching for candidate dots we make use of information provided by the previous frame, e.g. the search starts at the position where the dots were found in the last frame and we derive expected dot size characteristics from the last frame.

Edge Detection: A basic one dimensional convolution filter is used as an edge detector to identify light to dark and dark to light transitions while traversing the image scan lines horizontally. If a light to dark transition is followed by a dark to light transition, and the number of consecutive dark pixels is within range of the expected dot size, we might have encountered a candidate dot. Once it has been determined that in a minimum number of (vertically) neighboring scan lines a similar number of consecutive dark pixels is present the region is considered an actual candidate and passed on to the flood fill procedure.

Flood Fill: Starting at the darkest midpoint pixel of the neighboring dark pixels found by the edge detector, a local threshold flood fill is performed to determine the pixel memberships of the dark blob. The flood fill is aborted once it is determined that the dark region will not be classified as a candidate blob, for instance when it becomes too large, or when it hits the image border.

Ellipsoid Check: Since the dots in the marker pattern are circular, the candidate dots in the image will have to have ellipsoid characteristics. Once a region has been found that might be a candidate dot, a number of simple checks are performed to discard faulty candidates. These tests include checking for minimum and maximum width, height, and surface area, checking that the center of mass is approximately equal to the center of the bounding box, and that the center of mass should be located approximately midway the width and height of the region at the center of mass.

3.2.2. Candidate Matching, 3D Re-projection, and Pattern Search

Candidate matching and 3D re-projection are performed using standard epipolar geometry [4]. A candidate dot position is first transformed from image coordinates to camera coordinates, then corrected for lens distortion, and transformed into a rectified camera coordinate system used for both cameras. Next it is determined which candidate positions in the other image lie approximately on the same epipolar line, and the 3D positions of these matches are computed and added to the pool.

Each time new candidate dots and the corresponding 3D positions are added, it is checked whether 3D pattern can be found using these new positions. During the search it is ensured that no dot position is used twice. Furthermore, since we know the topology of the pattern, a number of permutations can be discarded.

For instance, we know that the dots will have to be arranged (counter-) clockwise in both images.

3.2.3. Synchronization

The iBOT FireWire cameras are not synchronized and exhibit some temporal drift. If this is not corrected, incorrect head positions might be determined or the 3D pattern might not be found at all. Since the images obtained from the cameras are marked with a timestamp, there are basically two ways to overcome the asynchronousness: interpolate the dot positions of the youngest image to estimate their positions at the time the oldest image was taken, or extrapolate the dot positions of the oldest image to estimate their positions at the time the youngest image was taken. Interpolating the youngest image's dots probably gives the most reliable results, but increases the delay. Extrapolating the oldest image's dots will not increase the delay but an overshoot might occur in the extrapolated dot positions. We chose the conservative approach and interpolated the dots of the youngest image.

3.3. Calibration

In any optical tracking system, the internal and external camera parameters have to be determined to be able to generate absolute 3D position and orientation parameters. For our head tracking system, we have adapted a calibration method developed by Zhang [29] which is provided in the Intel OpenCV image processing library [17]. We are currently working on a more automated calibration procedure.

4. Evaluation

We have implemented our tracking system on a 1.2 Ghz Athlon processor system with 512 Mb RAM, and used it in one of our PSS prototypes which is also equipped with a Logitech acoustical tracker. The cameras are mounted to the PSS chassis in front of the user, about 0.45 m apart. Figure 5 depicts the two iBOT cameras mounted next to the Logitech emitter.

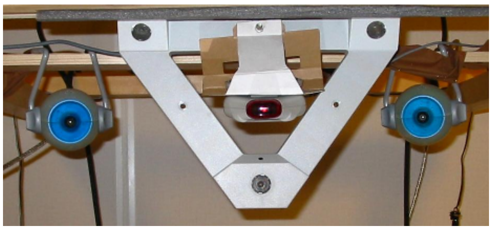

Figure 5: *The two iBOT cameras mounted in a PSS prototype next to the Logitech emitter.*

We have not implemented any additional filtering or prediction algorithms so as to evaluate both tracking systems on their bare results. Adding such filtering and prediction algorithms might improve the tracker performance when it will be put into actual use [8]. For simplicity we only compare the positional data of the two trackers.

4.1. Latency

To compare the delay of our optical tracking system to that of the Logitech acoustical tracking system we attached the acoustical receiver and the dot pattern together. These were freely moved inside the tracked volume. The output of both tracking systems was logged along with timestamps of when the results were available. This provided us with two tracks of positional data in time, one for each tracking system. A RMS best match was determined between the two tracks by shifting one track back or forth in time. This gives us the difference in delay between the two tracking systems.

The manual of the Logitech acoustical head tracker specifies a latency of 30 ms (plus an additional 20 ms for each additional filtering step required, which we did not; we ran both trackers without any filtering). The tracking delay difference between the optical tracker and the acoustical tracker was found to be 36 ms. Therefore, our optical tracking system has a delay of $30 + 36 = 66$ ms.

4.2. Accuracy

4.2.1. Method

Determining the absolute accuracy of any tracking system over the entire tracking volume is a tedious and very time-consuming process. A grid of sufficient resolution covering the tracking volume has to be determined. Next, the sensor or pattern has to be positioned accurately at each grid position, and the resulting tracker measurement has to be determined.

We have used a different approach. We positioned a flat plane inside the tracking volume at a number of different positions and orientations. We then moved the 3D pattern freely over this plane while logging the tracker results. Thus, the positions in one set of measurements should lie in one plane. We fitted an imaginary plane through the obtained measurements minimizing the RMS distance from the measurements to the plane. Although this procedure does not provide us with an absolute accuracy, we do get relative accuracy in the distances from the measured positions to the fitted plane.

Planes	Distances average	maximum	90%	99%
YZ	0.25	2.09	0.49	1.28
XY	0.69	3.44	1.18	2.02
XZ	0.16	0.62	0.32	0.49

Table 1: *Optical tracker measurement-to-plane distances in mm.*

Planes	Distances average	maximum	90%	99%
YZ	1.13	3.64	2.08	3.26
XY	3.26	109.4	12.61	20.38
XZ	2.06	6.05	3.90	4.82

Table 2: *Acoustical tracker measurement-to-plane distances.*

4.2.2. Results

Figure 6 shows the obtained measurement tracks for three planes: when moving the head up and down and forward and backward (YZ-plane), when moving up and down and left and right (XY-plane), and when moving left and right and forward and backward (XZ-plane), each for both tracking systems. Tables 1 and 2 summarize the obtained results. They depict the average distances from the obtained positions to the plane, the maximum distances, and the maximum distances of the 90% and 99% points closest to the plane.

From these tables it can be derived that the optical tracker performs very good and better than the Logitech acoustical tracker. In the YZ and XZ planes, 90% of the positions reported by the optical tracker are within sub-millimeter distance of the respective planes. Both trackers perform worst in the XY plane. For the optical tracker this is not surprising since depth estimations are known to be less accurate in narrow based stereo vision. The manual of the Logitech tracker does not state the exact methods used to derive the 3D positional data from the acoustic signals, so we cannot derive why it performs less in the XY plane compared to the other two planes.

The manual of the Logitech acoustical tracking system does specify a maximum positional error of 2% of the distance from the emitter to the receiver. In our measurements we regularly encountered higher error percentages than that. For instance, the distance from the XY plane to the transmitter was about 0.55 m, so a maximum error of 11 mm was to be expected. We obtained an average error of 3.26 mm which is about 0.6% of the receiver-transmitter distance, but 10% of the obtained measurements exhibited a distance larger than 12.61 mm to the plane, more than the specified 2%. The worst measurement exhibited an error of almost 20%.

An interesting phenomena revealed itself when analizing the obtained data. The XY plane measurement set for instance, revealed that when the acoustical tracker receiver was moved from right to left all tracker positions were located in front of the plane. When moved from left to right all positions were located behind the plane. Apparently, whereas our optical tracking system compensates for the asynchronousness of the cameras, the Logitech acoustical tracker does not compensate for the asynchronousness of the different emitters and receivers.

4.3. Update Rate, Robustness, and Computational Load

Figure 7 shows a log of tracker positions during 1.5 minutes of an interactive session. The track shows the user's 3D head positions in the PSS while he is working with an application. For clarity, their 2D projections are also depicted.

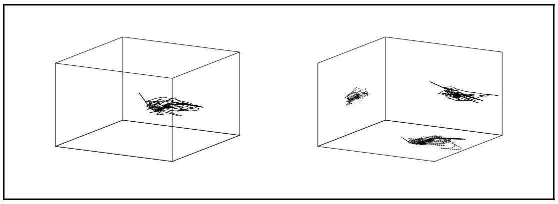

Figure 7: *Head positions during an interactive session. Left the 3D positions, right the 2D projections of the positions.*

Figure 7 clearly shows the limited head movements of the user. During the session, the tracker had a success rate of 100 %, and an update rate of 30 updates per second. Figure 8 shows a plot of the computational load put on the system by the tracking algorithm during the session. The average load is about 20%.

We performed another measurement where we moved the dot pattern vigorously (unrealistically) through the tracked volume. Such vigorous movements cause severe motion blur so the dot pattern cannot be found. Figure 9 plots the hit rate, and the computational load put on the system by the tracking algorithm. Clearly, the computational load is directly related to the hit rate. The computational load rises to 80% when the glasses cannot be found, and drops back to 20% when the hit rate rises to 100%.

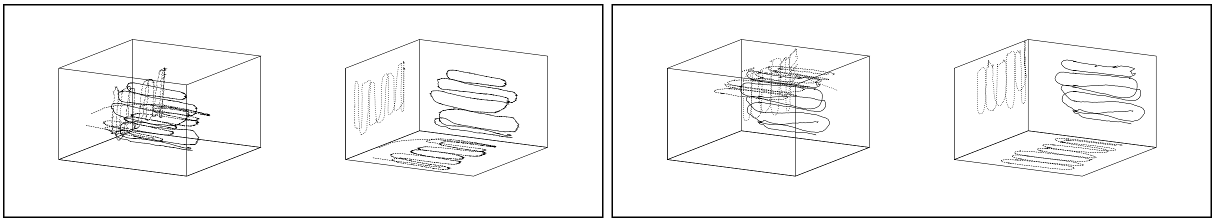

Figure 6: *Measurement tracks in 3 planes (YZ, XY, and XZ) inside the tracking volume of 0.5 m wide (X), 0.3 m high (Y), and 0.4 m deep (Z). Depicted are the 3D tracks in the tracking volume and 2D projections of the individual plane tracks. Left for the optical tracking system and right for the acoustical tracker.*

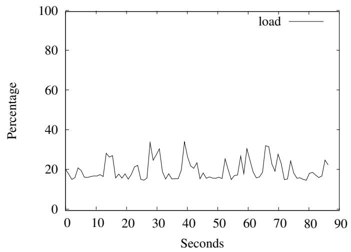

Figure 8: *Computational load during the interactive session of Figure 7.*

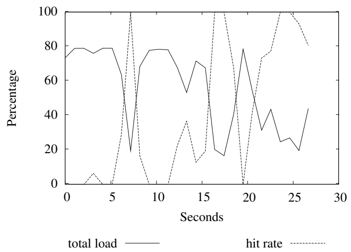

Figure 9: *Computational load during unrealistic head movements.*

The logitech acoustical tracking system does not put any computational load on the system since all calculations are performed in a separate device. The update rate of the Logitech tracker is 50 updates per second, as is stated in the manual.

5. Conclusion and Future Work

In this paper we have presented an affordable optical head tracking system for desktop-type VR/AR environments, based on commodity FireWire cameras. It provides accurate head tracking at 30 updates per second with a low computational load. Furthermore, the tracker is wire-less, non-intrusive, and interference free.

We have shown that the tracker is more accurate than the Logitech acoustical tracking system, while costing 10 times less. However, the optical tracker has longer latencies, caused by the maximum frame rate of 30 frames per second of the cameras. Since head tracking latencies in desktop-type VR/AR environments are less problematic than latencies in the tracking of interaction devices [25], we do recommend the tracker for head tracking, but doubt its use for interaction tracking.

Future research areas include the development of a (more) automated calibration procedure and the application of advanced filtering and prediction algorithms to lower the disadvantage of the longer latencies.

References

1. M. Andiel, S. Hentschke, T. Elle, and E. Fuchs. Eye-tracking for autostereoscopic displays using web cams. In *Proceedings of SPIE Stereoscopic Displays and Applications '02*, pages 78–86, 2002.

2. K.W. Arthur, K.S. Booth, and C. Ware. Evaluating 3D task performance for fish tank virtual worlds. *ACM Transactions on Information Systems*, 11(3):239–265, July 1993.

3. Y. Baillot and J. Rolland. Fundamental principles of tracking technology for virtual environments. Technical Report TR96–004, Center for Research and Education in Optics and Lasers (CREOL), Orlando Florida, November 1996.

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

5. S. Birchfield. Elliptical head tracking using intensity gradients and color histograms. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 232–237, 1998.

6. F.P. Brooks. What's real about virtual reality? *IEEE Computer Graphics and Applications*, 19(6):16–27, Nov./Dec. 1999.

7. C. Cruz-Neira, D.J. Sandin, and T.A. DeFanti. Surround-screen projection-based virtual reality: The design and implementation of the CAVE. In *Computer Graphics (SIGGRAPH '93 Proceedings)*, volume 27, pages 135–142, 1993.

8. M. Deering. High resolution virtual reality. In E.E. Catmull, editor, *Computer Graphics (SIGGRAPH '92 Proceedings)*, volume 26, pages 195–202, 1992.

9. K. Dorfmüller. An optical tracking system for VR/AR-applications. In M. Gervautz, A. Hildebrand, and D. Schmalstieg, editors, *Virtual Environments '99, Proceedings of the Virtual Environments Conference & fifth Eurographics Workshop*, pages 33–42, 1999.

10. D.O. Gorodnichy, S. Malik, and G. Roth. Affordable 3D face tracking using projective vision. In *Proceedings of International Conference on Vision Interface (VI'2002)*, pages 383–390, 2002.

11. D.O. Gorodnichy, S. Malik, and G. Roth. Nouse 'use your nose as a mouse' - a new technology for hands-free games and interfaces. In *Proceedings of International Conference on Vision Interface (VI'2002)*, pages 354–361, 2002.

12. M. La Cascia, S. Sclaroff, and V. Athitsos. 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):322–336, April 1999.

13. R. van Liere and J.D. Mulder. Optical tracking using projective invariant marker pattern properties. Accepted for Publication in the Proceedings of the IEEE Virtual Reality 2003 Conference, 2003.

14. L.-P. Morency, A. Rahimi, N. Checka, and T. Darrell. Fast stereo-based head tracking for interactive environments. In *Proceedings of Conference on Automatic Face and Gesture Recognition*, pages 375–380, 2002.

15. J.D. Mulder. The joint space station: A modular environment for collaborative virtual and augmented reality with a shared workspace. Submitted for Publication, 2003.

16. J.D. Mulder and R. van Liere. The personal space station: Bringing interaction within reach. In S. Richer, P. Richard, and B. Taravel, editors, *Proceedings of the Virtual Reality International Conference, VRIC 2002*, pages 73–81, 2002.

17. http://www.intel.com/research/mrl/research/opencv.

18. http://www.orangemicro.com.

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

20. J. Rekimoto. A vision-based head tracker for fish tank virtual reality: VR without head gear. In *Virtual Reality Annual International Symposium (VRAIS '95)*, pages 94–100, 1995.

21. M. Ribo. State of the art report on optical tracking. Technical Report VRVis 2001-25, TU Wien, 2001.

22. 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.

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

24. J. Ström, T. Jebara, S. Basu, and A. Pentland. Real time tracking and modeling of faces: An ekf-based analysis by synthesis approach. In *International Conference on Computer Vision: Workshop on Modelling People*, pages 55–61, 1999.

25. C. Ware and R. Balekrishnan. Reaching for objects in VR displays: Lag and frame rate. *ACM Transactions on Computer-Human Interaction*, 1(4):331–356, 1994.

26. C. Ware and G. Franck. Evaluating stereo and motion cues for visualizing information nets in three dimensions. *ACM Transactions on Graphics*, 15(2):121–140, April 1996.

27. G. Welch and E. Foxlin. Motion tracking: No silver bullet, but a respectable arsenal. *IEEE Computer Graphics and Applications*, 22(6):24–38, November/December 2002.

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

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

30. Z. Zivkovic and F. van der Heijden. A stabilized adaptive appearance changes model for 3D head tracking. In *IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems (RATFG-RTS'01)*, pages 175–182, 2001.