Tracking a head-mounted display in a room-sized environment with head-mounted cameras

Jih-fang Wang, Ronald Azuma, Gary Bishop
Vernon Chi, John Eyles, Henry Fuchs

Department of Computer Science
University of North Carolina
Chapel Hill, NC 27599-3175

ABSTRACT

This paper presents our efforts to accurately track a Head-Mounted Display (HMD) in a large environment. We review our current benchtop prototype (introduced in [WCF90]), then describe our plans for building the full-scale system. Both systems use an *inside-out* optical tracking scheme, where lateral-effect photodiodes mounted on the user's helmet view flashing infrared beacons placed in the environment. Church's method uses the measured 2D image positions and the known 3D beacon locations to recover the 3D position and orientation of the helmet in real-time. We discuss the implementation and performance of the benchtop prototype. The full-scale system design includes ceiling panels that hold the infrared beacons and a new sensor arrangement of two photodiodes with holographic lenses. In the full-scale system, the user can walk almost anywhere under the grid of ceiling panels, making the working volume nearly as large as the room.

1 INTRODUCTION

Many tracking systems developed for Head-Mounted Displays (HMDs) assume the user's body stays within a small space, such as inside a cockpit. At UNC, we are also interested in applications that let a user walk inside a large room while wearing a HMD. With our HMD system [Hol87][CHB$^+$89], we are developing "virtual worlds" where the user feels immersed inside a large, virtual environment that the user can explore and manipulate with hand-operated controls. For example, one visualization program lets users walk around a large virtual model of a molecule and dock a different drug molecule to it.

However, our current tracking device, the Polhemus [Pol80], does not offer satisfactory performance for these applications. The Polhemus suffers from a limited working volume, slow update rate, and interference from magnetic perturbations in the environment. The goal of our research is to overcome these limitations by designing and constructing an optical tracker that provides a large working range, high accuracy in the estimated head position, fast update rate, and immunity to electromagnetic interference. In this paper, we review a benchtop prototype of this system, which was described in more detail in [WCF90], then introduce our design of the full-scale system.

The remainder of this paper is organized as follows: Section 2 categorizes and surveys examples of current tracking schemes. Section 3 discusses the fundamental working principles of our tracker. Then in section 4, we briefly describe Church's method, which recovers the 3D position and orientation of the HMD. The errors in this method are analyzed in section 5, from which we discuss how to design the system to optimize performance. Section 6 briefly describes the implementation of the benchtop prototype and discusses the measured performance. In section 7, we introduce our design of the full-scale system. Finally, sections 8 and 9 offer conclusions and future directions to explore.

2 BACKGROUND

In this section, we survey existing approaches to 3D tracking. Commercial and experimental 3D position tracking devices fall into four categories: acoustic, mechanical, magnetic, and optical tracking. The first two methods, acoustic [Rob66] and mechanical [Kil76][Nol71][Vic74], are not suitable for our purposes. Because the speed of sound varies as the ambient air density changes, acoustic systems yield poor accuracy over long ranges. Acoustic systems also cannot sense orientation directly and are limited in bandwidth by the speed of sound. Mechanical tracking greatly restricts the range of motion of the user, yielding highly constrained working volumes.

The third category, magnetic tracking, includes a common tracking system in HMD systems: the Polhemus 3D position tracker [Pol80]. The Polhemus consists of two devices: a source and a sensor. The source generates a low-frequency magnetic field, and the sensor detects the polarization and orientation of the field. With this information, the Polhemus can calculate the position and orientation of the sensor relative to that of the radiating source.

The Polhemus offers two main advantages and several disadvantages. The two advantages are 1) the sensor attached to the user is small and lightweight, and 2) the sensor does not need to maintain an optical line-of-sight to the source. The Polhemus will still work even if the source and sensor are separated by some opaque, non-magnetic substance, like a human body. However, the Polhemus' disadvantages include a small working range, jittering, slow update rate, and sensitivity to magnetic materials in the environment. Any conducting materials or radiating sources present in the environment degrades the Polhemus' accuracy. In our Graphics Laboratory, the metal floor can curve the Polhemus' responses by as much as 30° near its spatial range, according to [CHB+89]. The Polhemus is not accurate enough (6 mm in position and 0.7° in rotation [Wan90]) to prevent "jittering." That is, when the user stands completely still, the virtual world appears to "swim" in front of the user because of the noise in the position and orientation readings. Experiments conducted in our HMD project also indicate that the Polhemus has a spherical working volume of about one meter in radius and provides only about 16 position updates per second.

The slow update rate causes a serious condition known as the lag or latency problem. When the user turns his or her head, the Polhemus takes up to 120 milliseconds to register and report the new position and orientation. In that time the user may have turned even further, so the images displayed no longer correspond with the direction the user is currently looking, seriously compromising the illusion of the virtual environment we are trying to generate. In severe cases, the lag can even induce motion sickness.

This leaves the fourth method, optical tracking, in which optical sensors (cameras, photodiodes) attempt to track optical sources (lights, beacons). Optical tracking reverses many of the advantages and disadvantages of magnetic tracking. It requires an unobstructed line-of-sight between multiple sources and sensors, but it works at long ranges, can be made fast and accurate, and is relatively immune to environmental distortions. We now survey several commercial and experimental systems that use optical tracking.

SELSPOT [Wol74] and OP-EYE [Uni81] are two commercial systems that use camera-like units with lateral-effect photodiodes as the detecting surfaces. These systems detect a single light source that shines on the surface of the photodiode. The 2D location where the light beam strikes the photodiode surface can be measured in real-time. A pair of these cameras can be used to measure the 3D location of the light source using stereopsis. The SELSPOT system is expensive ($40,000) and does not report the 3D position in real-time. The OP-EYE system has poor resolution and a very limited working volume [Wan90].

OPTOTRAK [Nor88] uses one camera with two dual-axis CCD infrared position sensors. Each position sensor has a dedicated processor board to calculate the image position of the light source. Again, the triangulation principle is applied to recover the position of the light source in space. The system is expensive, and the bulky camera weighs more than 10 pounds.

3 THE INSIDE-OUT OPTICAL METHOD

None of the surveyed systems provides sufficient performance to achieve our goals. Because optical tracking seems to possess the most potential for our purposes, we have been working on a system based on optical tracking methods.

Most commercially available optical trackers use a scheme we call *outside-in* tracking. This approach places the sensors at known, fixed locations in the environment and attaches beacons to the user's helmet. The sensors attempt to track these beacons as the user walks around the room. The outside-in approach, although intuitively simple and appealing, places stringent demands on the sensor's resolution. For example, consider a room 4 meters wide on each side. We mount a camera on one side of the room and try to track a 0.1° head rotation at the opposite side of the room. The 0.1° rotation will move a light source attached to the user's head by less than 1 mm. To detect a 1 mm movement of the light source, a sensor must be capable of resolving 1 part in 4000 (1 mm / 4 m). This requirement is difficult to meet by either CCD sensors or lateral-effect photodiodes.

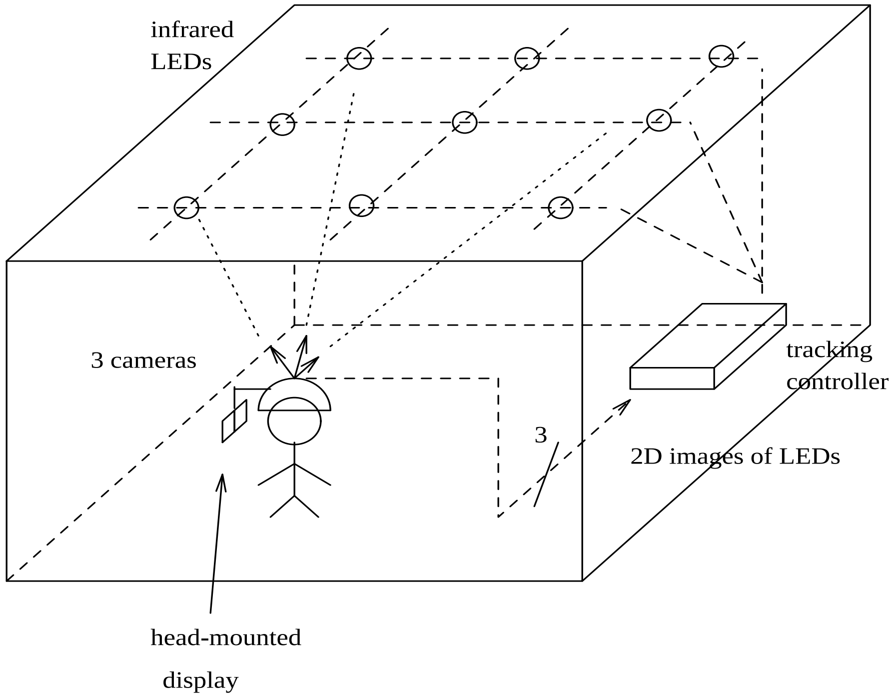

Figure 1: The inside-out tracking system

In our system, we reverse this outside-in configuration and use an *inside-out* tracking scheme, in which we mount the optical sensors on the helmet and place beacons on the ceiling of the room. With a dense pattern of beacons on the ceiling, each camera needs to cover only a small patch of the ceiling to insure that it sees a beacon. With such a small viewing field, a slight rotation of the user's head causes a large shift of the beacon's position in the sensor image. Hence, a sensor with relatively low resolution can detect the same 0.1° rotation. For example, if a camera uses a 50 mm lens and a detection surface of 1 cm^2, then the field of view of the camera at 4 meters away is only 0.8 x 0.8 m^2. A camera that can resolve 1 part in 800 can detect a 1 mm movement easily.

Our system uses lateral-effect photodiodes as the optical sensors, which offer several advantages over CCD arrays. A lateral-effect photodiode is a large photosensitive surface, usually square in shape, which measures the x and y location of the centroid of a luminous spot. Lateral-effect photodiodes provide faster response and higher positional resolution than CCD arrays, because it is easier to calculate the location of a centroid with the former than the latter. Furthermore, photodiodes have no dead-zone over the recording surfaces and provide useful positional readings even if the light spot is out of focus or blurred, unlike CCD arrays.

For beacons, our system uses infrared light emitting diodes (LEDs) that have high output power and wide emission angles. The LEDs operate in the infrared spectrum so the user will not be distracted by their constant blinking. Infrared filters placed over our photodiodes let us use ordinary fluorescent lamps to light the room without interfering with the beacons. Figure 1 depicts the configuration of this inside-out tracking system.

4 ALGORITHMS FOR INFERRING 3D POSITION

We now describe a method for recovering the position and orientation of the user's head. Our system does this by trying to match the 3D positions of the observed beacons with their 2D images on the photodiode surfaces. Here we describe a method proposed by Earl Church [Chu45], first used in aerial photogrammetry: the process of finding where a camera was when it took an aerial photograph by locating three known landmarks in the photograph. Although we will be using photodiodes instead of cameras in our system, Church's discussion refers to the optical sensors as "cameras," so we will use that term while describing Church's method.

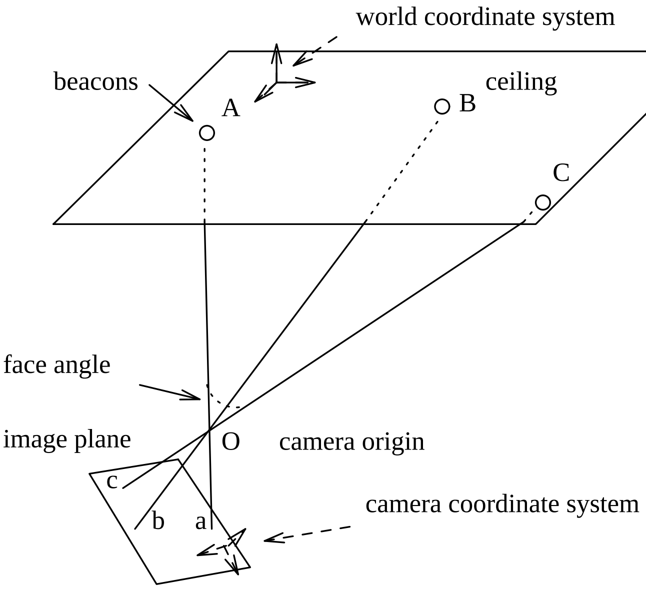

Figure 2: Church's algorithm

Figure 2 shows how we apply Church's method to our inside-out tracking system. We define two coordinate systems: the world coordinate system, which is aligned to the fixed room environment, and the camera coordinate system, which uses the nodal point of the camera as the origin and the image plane of the camera as the x-y plane. Church's method assumes our camera sees three beacons. Any two of the three observed beacons form a *face angle* with the camera origin. Church's solution is based on the condition that the face angle subtended by any two beacons in space is equal to the face angle subtended at their corresponding image locations. The face angles formed by the camera origin and the beacon locations in the image plane (points a, b, and c in Figure 2) can be calculated directly in the camera coordinate system. However, because the location of the camera nodal point in the world coordinate system is unknown, the face angles subtended by the beacons in space are unspecified. Church's method starts by hypothesizing the position of the camera origin in the world coordinate system. Usually we use the last known camera

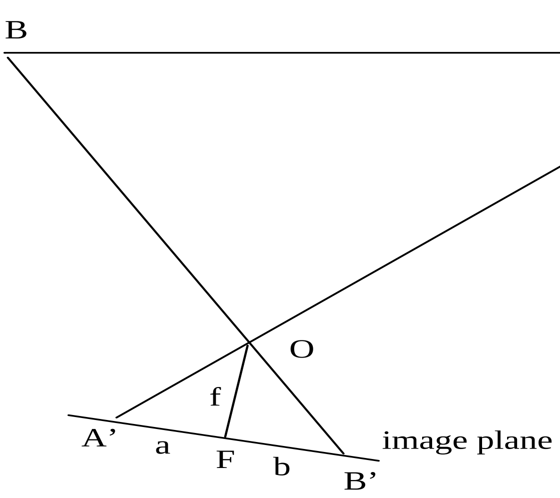

Figure 3: Error on the image plane

position as the hypothesis. With this guess, we can calculate the three face angles in the world coordinate system. In general, these face angles will not match those computed from the image unless the hypothesized camera position is correct. An iterative search scheme is employed to locate the correct 3D camera location by minimizing the difference between the corresponding face angles.

5 SYSTEM ERROR

The system error comes from errors in measuring the two sets of face angles that Church's method tries to match. The error of the face angles measured in camera coordinates is mainly due to the camera's resolution limits, and can be expressed by [WCF90]

$$\varepsilon_\theta = (\frac{1}{a^2+f^2} + \frac{1}{b^2+f^2})\frac{fD}{2r}, \tag{1}$$

where D is the width of the square photosensitive surface, r is the resolution of the photodiode, and a, f, and b are shown in Figure 3. To minimize the error in Equation 1, we should make the resolution of the camera as high as possible, use a lens with a long focal length, and keep the separation of the beacon positions in the image plane as large as possible.

The error of the face angles measured in world coordinates is mainly due to the uncertainty of the positions of the beacons we flashed, and can be expressed as

$$\varepsilon_\theta = \left(\frac{x_a+y_a}{|OA|^2} + \frac{x_b+y_b}{|OB|^2} \right) \varepsilon_p, \tag{2}$$

where the coordinates of A are (x_a, y_a), the coordinates of B are (x_b, y_b), and ε_p bounds the beacon placement error. The error in Equation 2 is directly proportional to ε_p, and increasing the separation of the beacon positions in the image plane decreases this error.

Equation 1 and Equation 2 conflict on how to minimize the error in the system. Equation 1 states that we need a lens with a long focal length to reduce the area the camera sees, making it easier to track a small amount of movement within that area. But a small viewing field decreases the separation of the beacon positions in the image plane, thus increasing the error of the face angle measured in world coordinates, according to Equation 2. We avoid this conflict by using several cameras observing widely separated directions, rather than using just one camera. With multiple cameras, each camera observes a small area to satsify Equation 1, while the wide separation of view directions satisfies Equation 2.

6 THE BENCHTOP PROTOTYPE

We have constructed a desktop prototype system to prove the correctness of our design. Although limited in performance, this prototype nonetheless demonstrates the integration and coordination of all essential components of our system. Three identical photodiodes (Figure 5), pointed in widely separated directions (Figure 6), are mounted on a stand that we can manually rotate and translate (Figure 7). Three infrared LED beacons are mounted on the ceiling, with one LED in each photodiode's field of view. This prototype system provides about 16° of rotational freedom and 1.5 feet of translational freedom.

The other elements of the prototype are a host computer (μVax II), a controller which directs the data acquisition from the tracker, and signal processing circuitry. Signals are sent to and from the host computer through a DRV-11 parallel interface. We now outline the steps the system takes during each iteration. First, the host sends a *go* signal to the controller, which in turn directs the signal processing circuitry to integrate the photodiode signals for 100 μsec with no LEDs lit. This "dark current" measurement improves the resolution of the photodiode and helps reduce the noise inherent in the photodiode readout. Then the host computer reads this measurement and sends another *go* signal to the controller. Now the controller pulses the LEDs for 100 μsec, integrates the signals, and sends them to the host computer as the "light" measurement. The host subtracts the "dark" measurement from the "light" measurement and uses this result to compute the X and Y locations of the beacons in each photodiode's image plane. Finally, the host uses Church's method to recover the 3D position and orientation of the unit. An overview of the system is shown in Figure 4.

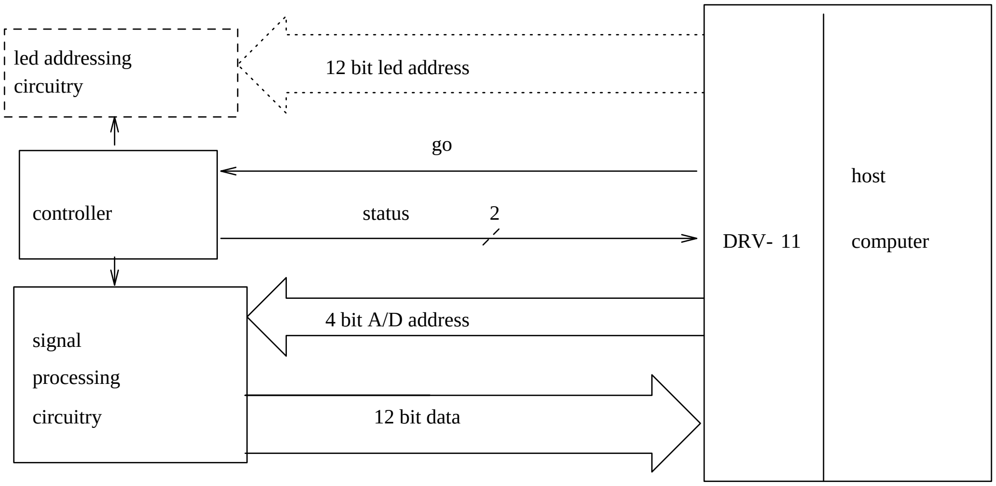

Figure 4: Prototype system configuration

We quantitatively measured the accuracy, range, and speed of this prototype [WCF90]. From the experiments conducted, we estimate that this design offers excellent accuracy in tracking head motion (0.1° in rotation and 2 mm in translation) at long range (3 meters between source and sensor). Because of this accuracy, we have observed almost no jittering in our prototype. Although the photodiodes and signal processing circuitry can run at 1000 Hz, the current prototype only provides about 25 updates/sec, because of the time it takes our slow μVax II host to compute Church's method. Simulated runs on a faster machine, such as a DECstation 3100, yield a fast update rate ($\sim$ 200 updates/sec), which we expect to achieve in the full system. A fast host computer will also reduce the lag to about 5 milliseconds.

Figure 5: One of the photodiodes

Figure 6: The configuration of three photodiodes

Figure 7: The benchtop system

7 DESIGN OF THE FULL-SCALE SYSTEM

We now discuss our design of a full-scale system, based on this prototype, that will be capable of replacing the Polhemus in our HMD system. Besides changing to a faster workstation, our design includes a method of mounting large numbers of beacons in the environment and a new configuration of sensors to be placed on the user's helmet.

The first change is to create a dense array of beacons covering the ceiling of a room. We must have many beacons in the full-scale system because each photodiode has a narrow field of view, and Church's method requires that we see three beacons at all times. We chose to place the beacons on the ceiling, rather than on the walls, because the user's helmet will have an unobstructed view of the ceiling, whereas obstacles may prevent it from seeing parts of the walls. Since our ceiling is composed of 2' x 2' easily-removable ceiling tiles, we will build 2' x 2' panels to replace these tiles. Each panel will have 16 infrared LEDs, arranged in a hexagonal grid pattern, to minimize the maximum distance between any two neighboring LEDs (Figure 8). Each panel also includes power transistors to light the LEDs, connectors that hook the panels together in one long daisy chain, and address decoding circuitry that lets our controller light any single LED in the entire array of panels by providing one 12-bit address. We intend to handwire three panels to replace the three LEDs in our current prototype, providing greater rotational and translational freedom while also testing our panel design. For the full system of approximately 100 to 150 panels, we will design and use large printed circuit boards which have the components soldered onto them.

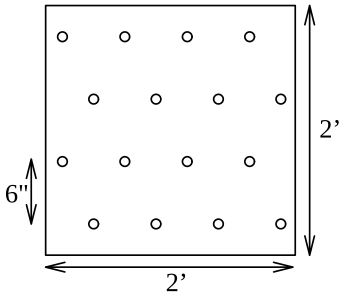

Figure 8: Panel with 16 LEDs in hexagonal grid pattern

By placing the beacons only on the ceiling, it is possible to create a system where the working volume is almost as large as the room. To do this, we require that the photodiodes on the user's helmet rely only on the beacons in a small patch of ceiling directly above the user's head. If we can accomplish this, we can accurately track the user just about anywhere underneath the grid of panels, creating a large working volume that the user can move within. With enough panels, this volume can be almost as large as the room itself.

But to make this idea work, we must replace the prototype configuration of three photodiodes. The only way to rely solely on the patch of ceiling directly above the user's head is to add more photodiodes to the helmet, since helmet pitch and roll (but not yaw) can rotate a vertical photodiode away from the ceiling and toward a wall instead. We arrange the extra photodiodes so that as the helmet starts rotating some photodiodes away from the ceiling, others will rotate toward the ceiling, so three photodiodes will always see beacons above us. We wrote some simulation programs to test many possible configurations and found some that worked with eight photodiodes.

Unfortunately, that many photodiodes would be uncomfortably heavy if placed on the user's head, so we avoid the weight problem by using holographic lenses. A holographic lens can superimpose several different view directions onto one photodiode. Taking into account the restrictions of the holographic lens, we designed a configuration of two photodiodes, each with a holographic lens that looks in four different directions (Figure 9), for a total of eight different views. The holographic lens is considerably lighter than the regular lens we are currently using, so the proposed configuration will actually be much lighter than our current unit of three photodiodes.

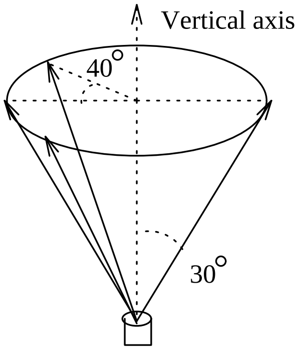

Figure 9: Holographic lens with four views

To say a configuration works, we must also assume some reasonable limits on how far the user can tilt his or her head. Certainly the user cannot do a handstand, or all the photodiodes will face the floor. During the search, we assumed that a user would want more freedom in tilting his or her head forward and backward than in tilting it from side-to-side. Figure 10 shows a mounting scheme that performed well. Both photodiodes use the holographic lens pattern shown in Figure 9, with the rear lens flipped around 180°. Figure 11 shows this configuration underneath a 3 x 3 array of panels, with the area each view sees on the ceiling outlined by a quadrilateral. This configuration lets the user rotate 360° around the vertical axis and tilt 70° forward and 60° back. Side-to-side tilt limits are a function of the front-to-back tilt, varying from ±37° to ±5°. These numbers represent a worst case, with the photodiodes depending only on the infrared beacons in a 3 foot radius around the point directly above the user's head, which may occur when the user walks within 3 feet of the edge of the ceiling array. In more typical positions near the center of the panel array, the photodiodes can rely on a much larger patch of ceiling, with correspondingly greater rotational freedom.

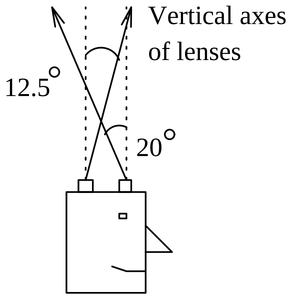

Figure 10: Mounting two holographic lenses on user's head

Will the full-scale system maintain the accuracy measured in the benchtop prototype? That depends on the energetics. We achieve the stated accuracy with our prototype system when a beacon is 3 meters away. Now when we place the photodiodes on the helmet, the beacons will be roughly one meter away from the photodiodes, depending on the user's height, which provides a gain in energetics due to the reduced distance. But the holographic lenses will cost us in energetics, since they superimpose N views onto one point (reducing the light per view by 1/N), and each view is tilted 30° off-axis (creating a cosine aperture loss). We

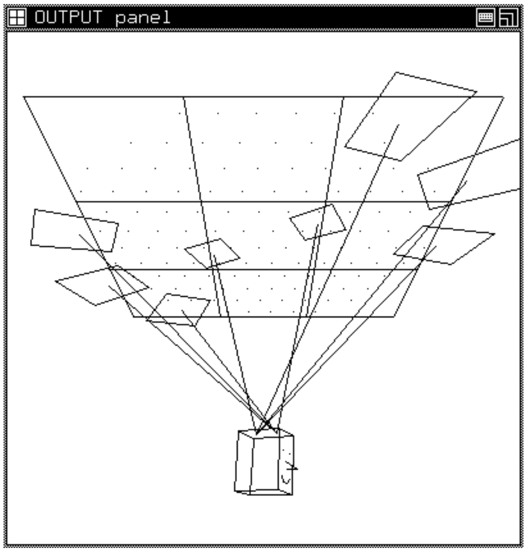

Figure 11: New configuration of eight photodiode views under a 3x3 array of panels

estimate that the break-even point occurs when the user is 5' 7". Taller users will enjoy somewhat greater accuracy, and shorter users will suffer somewhat reduced accuracy, but we do not expect the differences to be large.

8 CONCLUSIONS

We have constructed a benchtop prototype of an optical inside-out tracking system that offers significant advantages in speed, range, and accuracy over the Polhemus tracker and other commercially available systems. This proof-of-concept prototype will lead to a full-scale system that can track a user's head in a large environment with a high degree of accuracy and at rapid update rates, something that to our knowledge has not been achieved before. When completed, this system will replace our Polhemus headtracker.

9 FUTURE WORK

After finishing this system, we hope to explore ways to reduce the large number of LED beacons required in our system, perhaps by building hybrid systems that combine a non-optical tracking scheme with our optical tracker. For example, inertial methods would let us follow the user's head even when the photodiodes can't see three beacons. When the photodiodes occasionally do see enough beacons, we could use that measurement to reduce the accumulated drift inaccuracies from the non-optical measurements. Such a hybrid scheme would increase the robustness of the system and reduce the required density of beacons in our environment, reducing the cost and making it more portable. Ideally, "smart" optical sensors, such as the type Bishop investigated [Bis84], combined with inertial trackers may almost eliminate the need for optical beacons.

10 ACKNOWLEDGEMENTS

Thanks are due to Brad Bennett and John Thomas of the Microelectronics Systems Laboratory of UNC for their help with the system programming and hardware. Dr. John H. Halton developed an alternate method for recovering the 3D position and orientation of the HMD from three beacons. Our work is supported in part by ONR grant N00014-86-0680, NIH Division of Research Resources grant RR 02170-05, and a Pogue Fellowship.

11 REFERENCES

[Bis84] T. G. Bishop. *Self-Tracker: A Smart Optical Sensor on Silicon.* PhD thesis, U. of North Carolina, Chapel Hill, NC, 1984.

[CHB+89] J. C. Chung, M. R. Harris, F. P. Brooks Jr., H. Fuchs, M. T. Kelley, J. W. Hughes, M. Ouh-Young, C. Cheung, R. L. Holloway, and M. Pique. Exploring virtual worlds with head-mounted displays. In *Proceedings SPIE Conference, Nonholographic True Three-Dimensional Display Technologies*, Los Angeles, CA, Jan. 1989.

[Chu45] E. Church. Revised geometry of the aerial photograph. In *Bulletin of Aerial Photogrammetry*, volume 15. Syracuse University, 1945.

[FGH+85] H. J. Fuchs, J. Goldfeather, J. P. Hultquist, S. Spach, J. Austin, F. P. Brooks Jr., J. Eyles, and J. Poulton. Fast spheres, textures, transparencies, and image enhancements in pixel-planes. *Computer Graphics*, 19(3), 1985.

[Hol87] R. Holloway. Head-mounted display technical report. Technical Report 87-015, U. of North Carolina, 1987.

[Kil76] P. J. Kilpatrick. *The Use of a Kinesthetic Supplement in an Interactive Graphics System.* PhD thesis, U. of North Carolina, Chapel Hill, NC, 1976.

[Nol71] M. A. Noll. *Man-machine Tactile Communication.* PhD thesis, Polytechnic Institute of Brooklyn, Brooklyn, NY, 1971.

[Nor88] Northern Digital. *Trade literature on Optotrak - Northern Digital's Three Dimensional Optical Motion Tracking and Analysis System.* Northern Digital Inc., Waterloo, Ontario, Canada, 1988.

[Pol80] Polhemus Navigation Sciences Division. *3Space Isotrak User's Manual.* McDonnell Douglas Electronics Company, 1980.

[Rob66] L. G. Roberts. The lincoln wand. In *FJCC*, Washington, D.C., 1966. Spartan Books.

[Uni81] United Detector Technology. *Trade literature on OP-EYE optical position indicator.* United Detector Technology Inc., Santa Monica, CA, 1981.

[Vic74] D. L. Vickers. *Sorcerer's Apprentice: Head-mounted Display and Wand.* PhD thesis, U. of Utah, Salt Lake City, UT, 1974.

[Wan90] J. F. Wang. *A Real-time 6D Optical Tracker for Head-mounted Display Systems.* PhD thesis, U. of North Carolina, Chapel Hill, NC, 1990.

[WCF90] J. F. Wang, V. Chi, and H. Fuchs. A real-time 6D optical tracker for head-mounted display systems. In *1990 Symposium on Interactive 3D Graphics.* Snowbird, Utah, 1990.

[Wol74] H. J. Woltring. New possibilities for human motion studies by real-time light spot position measurement. *Bioelemetry*, 1, 1974.