The HiBall Tracker: High-Performance Wide-Area Tracking for Virtual and Augmented Environments

Greg Welch, Gary Bishop, Leandra Vicci, Stephen Brumback, Kurtis Keller, D'nardo Colucci†

University of North Carolina at Chapel Hill
Department of Computer Science, CB# 3175
Chapel Hill, NC 27599-3175 USA
01-919-962-1700
{welch, gb, vicci, brumback, keller}@cs.unc.edu

†Alternate Realities Corporation
27 Maple Place
Minneapolis, MN 55401 USA
01-612-616-9721
colucci@virtual-reality.com

1. ABSTRACT

Our HiBall Tracking System generates over 2000 head-pose estimates per second with less than one millisecond of latency, and less than 0.5 millimeters and 0.02 degrees of position and orientation noise, everywhere in a 4.5 by 8.5 meter room. The system is remarkably responsive and robust, enabling VR applications and experiments that previously would have been difficult or even impossible.

Previously we published descriptions of only the Kalman filter-based software approach that we call Single-Constraint-at-a-Time tracking. In this paper we describe the complete tracking system, including the novel optical, mechanical, electrical, and algorithmic aspects that enable the unparalleled performance.

1.1 Keywords

virtual environments, tracking, calibration, autocalibration, delay, latency, sensor fusion, Kalman filter, optical sensor

2. INTRODUCTION

In 1991 the University of North Carolina demonstrated a working scalable optoelectronic head-tracking system in the *Tomorrow's Realities* gallery at that year's ACM SIGGRAPH conference [24, 25, 26]. The system used four head-mounted lateral effect photo diode (LEPD) sensors that looked upward at a regular array of infrared light-emitting diodes (LEDs) installed in precisely machined ceiling panels as shown in Figure 1. Photogrammetric techniques were used to compute a user's head position and orientation using the known LED positions and their projected images on each LEPD sensor [4]. The system was ground-breaking because it was unaffected by ferromagnetic and conductive materials in the environment, and the working area of the system was determined solely by the number of ceiling panels. See the left panel in Figure 1, and color plate image Welch 1.

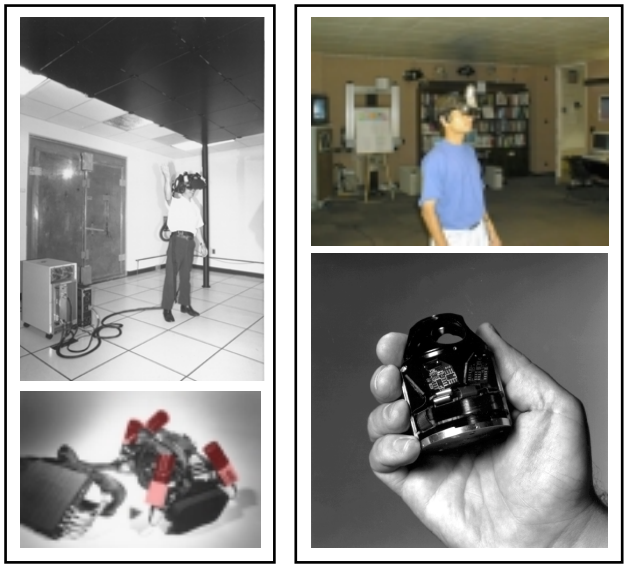


Figure 1. Left: the original UNC optoelectronic ceiling tracker in use, and a close-up of the head-mounted display and sensor fixture, along with the signal processing and communications electronics pack. Right: the new system in use, and a close-up of the self-contained HiBall with lenses and part of the cover removed. See also the color plate Welch 1.

In this paper we present a new and vastly improved version of that 1991 system. We call the new system the *HiBall Tracker.* Thanks to significant improvements in both hardware and software this new system offers unprecedented speed, resolution, accuracy, robustness, and flexibility. In particular, the bulky and heavy cameras and backpack of the previous system have been replaced by a small head-mounted *HiBall*. In addition, the precisely machined LED ceiling panels of the previous system have been replaced by lower-tolerance panels that are relatively inexpensive to make and simple to install. See the right panel in Figure 1, and color plate image Welch 1. Finally, we are using an unusual Kalman-filter-based approach to tracking that generates very accurate tracking estimates at a high rate with low latency, and simultaneously self-calibrates the system.

As a result of these improvements the HiBall Tracker can generate over 2000 estimates per second, with less than one millisecond of latency. The system exhibits sub-millimeter translation noise and similar measured accuracy, as well as less than 0.03 degrees of orientation noise with similar measured accuracy. The weight of the user-worn HiBall is about 300 grams, making it lighter than just *one camera* in the 1991 system. The working volume of the current system is greater than 90 cubic meters (greater than 45 square meters of floor space, greater than 2 meters of height variation). This area can be expanded by adding more tiles, or by using checkerboard configurations which spread tiles over a larger area.

3. SYSTEM OVERVIEW

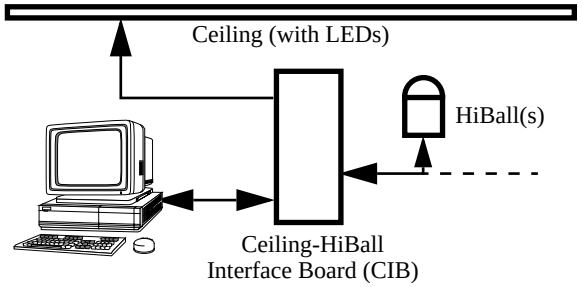


Figure 2. A block diagram of the HiBall tracking system.

The HiBall tracker system (Figure 2) provides six-degree-of freedom tracking of devices in real time. An outward-looking infrared-sensing subsystem called a HiBall (Figure 1, lower-right) is mechanically fixed to each device to be tracked. The HiBalls view an environment containing a subsystem of fixed-location infrared beacons which we call the Ceiling. At the present time, the beacons are in fact entirely located in the ceiling of our laboratory, but could as well be located in walls or other arbitrary fixed locations. These subsystems are coordinated by a Ceiling-HiBall Interface Board (CIB) which provides communication and synchronization functions between the host computer and the attached subsystems. Each HiBall has 26 narrow (less than 6 degree) views distributed over a large solid angle. Beacons are selectively flashed in a sequence such that they are seen by many different fields of view of each HiBall. Initial acquisition is performed using a brute force search through beacon space, but once initial lock is made, the selection of beacons to flash is tailored to the fields of view of the HiBalls. Tracking is maintained using a Kalman-filter-based prediction-correction algorithm known as SCAAT. This technique has been further extended to provide self-calibration of the Ceiling on-line with the tracking of the attached HiBalls.

4. SYSTEM COMPONENTS

4.1 The Ceiling

The Ceiling architecture provides for flashing only one beacon at a time. The beacons may be flashed in any sequence, but protection is provided in hardware and software to prevent exceeding the duty cycle of the infrared (IR) light emitting diodes (LEDs). Thus, no single LED can be flashed again until it has had sufficient time to cool. LED driving current and therefore emitted light level is selectable for use by a software automatic gain control (AGC) function as described in section 5.2.

Figure 3. Left: Bishop lifts some Ceiling panels. Right: an individual Ceiling panel module showing the main PC board with four LED strips (eight LEDs per strip).

As presently implemented, the beacons are packaged in modules, physically 61 centimeters square, to drop into a standard false ceiling grid (Figure 3 and color plate image Welch 2). Each module contains 32 LEDs in four strips, resulting in a rectangular pattern with periods of 7.6 and 15.2 centimeters, respectively. We currently have enough panels to cover an area approximately 5.5 by 8.5 meters, for a total of approximately 3,000 LEDs. The LEDs are Siemens SFH-487P GaAs diodes which provide both a wide angle radiation pattern and high peak power, emitting at a center wavelength of 880 nm in the near IR. These devices can be pulsed up to 2.0 Amps for a maximum duration of $200\mu s$ with a 1:50 (on:off) duty cycle.

The Ceiling panel modules are daisy-chain connected, with module selection encoding being position rather than device dependent. Operational commands are presented to the first module of the daisy chain. At each module, if the module select code is zero the module decodes and executes the operation; else it decrements the module select code and passes it along to the next module. Upon decoding, a particular LED is selected, a drive level is established, and the LED is flashed for up to $200\mu s$ (in $20\mu s$ increments).

4.2 The HiBall

As can be seen in Figure 1 and color plate image Welch 1 the HiBall is a hollow ball having dodecahedral symmetry with lenses in the upper six faces and lateral effect photo diodes (LEPDs) on the insides of the opposing six lower faces. This immediately gives six primary fields of view, or camera systems which share the same internal air space, and whose adjacent directions of view are uniformly separated by 57 degrees.

While the original intent of the shared internal air space was to save space, we subsequently realized that light entering any lens sufficiently off axis can be seen by an adjacent LEPD. As such, five secondary fields of view are provided by the top or central lens, and three secondary fields of view are provided by the five other lenses. Overall, this provides 26 fields of view which are used to sense widely separated groups of beacons in the environment. While these extra

views complicate the initialization of the Kalman filter as described in section 5.5, they turn out to be of great benefit during steady-state tracking by effectively increasing the overall HiBall field of view without sacrificing resolution.

The lenses are simple plano-convex fixed focus lenses. IR filtering is provided by fabricating the lenses themselves from RG-780 Schott glass filter material which is opaque to better than 0.001% for all visible wavelengths, and transmissive to better than 99% for IR wavelengths longer than 830 nm. The longwave filtering limit is provided by the DLS-4 LEPD silicon photodetector (UDT Sensors, Inc.) with peak responsivity at 950 nm but essentially blind above 1150 nm.

The LEPDs themselves are not imaging devices; rather they detect the centroid of the luminous flux incident on the detector. The x-position of the centroid determines the ratio of two output currents, while the y-position determines the ratio of two other output currents. The total output current of each pair are commensurate, and proportional to the total incident flux. Consequently, focus is not an issue, so the simple fixed-focus lenses work well over a range of beacon distances from about half a meter to infinity.

Finally, the LEPDs and associated analog and digital electronic components are mounted on a custom rigid-flex printed circuit board as shown in color plate image Welch 2. This arrangement makes efficient use of the internal HiBall volume while maintaining isolation between analog and digital circuitry, and increasing reliability by alleviating the need for inter-component mechanical connectors.

4.3 The Ceiling-HiBall Interface Board

The Ceiling-HiBall Interface Board (CIB), shown below in Figure 4, provides communication and synchronization between a host personal computer, the Ceiling (section 4.1) and the HiBall (section 4.2).

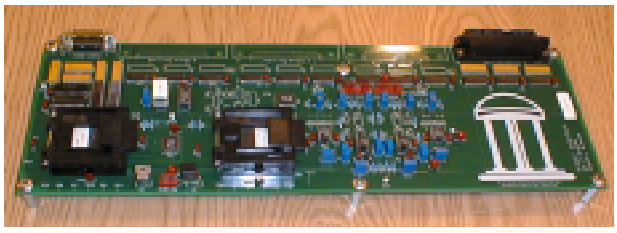


Figure 4. The Ceiling-HiBall Interface Board (CIB). The CIB shown is 19 inches, the newest revision is 14 inches.

The CIB has four Ceiling ports allowing interleaving of ceiling panels for up to four simultaneous led flashes and/or higher Ceiling bandwidth for more simultaneous hiball usage. (The Ceiling bandwidth is inherently limited by LED current restrictions as described in section 4.1, but this can be increased by spatially multiplexing the Ceiling tiles.) The CIB has two tether interfaces that can communicate with up to four daisy-chained hiballs each. The full-duplex communication with the hiballs uses a modulation scheme (BPSK) allowing future wireless operation. The interface from the CIB to the host PC is the stable IEEE1284C extended parallel port (EPP) standard.

The CIB comprises analog drive and receive components as well as digital logic components. The digital components implement store and forward in both directions and synchronize the timing of the LED “on” interval within the HiBall dark-light-dark intervals. The protocol supports full-duplex flow control. The data are arranged into packets containing error detection to insure data quality.

5. METHODS

5.1 Bench-Top (Off-Line) HiBall Calibration

After each HiBall is assembled we perform an off-line calibration to determine the correspondence between image-plane coordinates and rays in space. This involves more than just determining the view transform for each of the 26 camera (sensor) views. Non-linearities in the silicon sensor and distortions in the lens (e.g., spherical aberration) cause significant deviations from a simple pin-hole camera model. We dealt with all of these issues through the use of a two-part camera model. The first part is a standard pin-hole camera represented by a 3x4 matrix. The second part is a table mapping real image-plane coordinates to ideal image-plane coordinates.

Both parts of the camera model are determined using a calibration procedure that relies on a goniometer (an angular positioning system) of our own design. This device consists of two servo motors mounted together such that one motor provides rotation about the vertical axis while the second motor provides rotation about an axis orthogonal to vertical. An important characteristic of the goniometer is that the rotational axes of the two motors intersect at a point at the center of the HiBall optical sphere; this point is defined as the origin of the HiBall. (It is this origin that provides the reference for the HiBall state during run time as described in section 5.3.) The rotational positioning motors were rated to provide 20 arc-second precision; we further calibrated them using a surveying grade theodolite, an angle measuring system, to 6 arc seconds.

In order to determine the mapping between sensor image-plane coordinates and three-space rays, we use a single LED mounted at a fixed location in the laboratory such that it is centered in the view directly out of the top lens of the HiBall. This ray defines the Z or up axis for the HiBall coordinate system. We sample other rays by rotating the goniometer motors under computer control. We sample each view with rays spaced about every 6 minutes of arc throughout the field of view. We repeat each measurement 100 times in order to reduce the effects of noise on the individual measurements and to estimate the standard deviation of the measurements.

Given the tables of approximately 2500 measurements for each view, we first determine a 3 by 4 view matrix using standard linear least-squares techniques. Then we determine the deviation of each measured point from that predicted by the ideal linear model. These deviations are re-sampled into

a 25 by 25 grid indexed by sensor-plane coordinates using a simple scan conversion procedure and averaging. Given a measurement from a sensor at run time we convert it to an "ideal" measurement by subtracting a deviation bi-linearly interpolated from the nearest 4 entries in the table.

5.2 On-Line HiBall Measurements

Upon receiving a command from the CIB (section 4.3), which is synchronized with a CIB command to the ceiling, the HiBall selects the specified LEPD and performs three measurements, one before the beacon flashes, one during the beacon flash, and one after the beacon flash. Known as "dark-light-dark", this technique is used to subtract out DC bias, low frequency noise, and background light from the beacon signal.

Each LEPD has four transimpedance amplifiers, the analog outputs of which are multiplexed with those of the other LEPDs, then sampled, held, and converted by four 16-bit Delta-Sigma ADCs. Multiple samples can be integrated internally in the HiBall. The digitized LEPD data are organized into a packet for communication back to the CIB. The packets also contain information to assist in error-detection. The communication protocol is simple, and while presently implemented by wire, the modulation scheme is amenable to a wireless implementation. The present wired implementation allows multiple HiBalls to be daisy chained so a single cable can support a user with multiple HiBalls.

During run time we attempt to maximize the signal-to-noise ratio of the measurement with an automatic gain control scheme. For each LED we store a target signal strength constant. We compute the LED current and number of integrations (of successive A/D samples) by dividing this strength constant by the square of the distance to the LED, estimated from the current position estimate. After a reading we look at the strength of the actual measurement. If it is larger than expected we reduce the gain, if it is less than expected we increase the gain. The increase and decrease are implemented as on-line averages with scaling such that the gain constant decreases rapidly (to avoid overflow) and increases slowly. Finally we use the measured signal strength to estimate the noise on the signal using [8], and then use this as the measurement noise estimate for the Kalman filter (section 5.3).

5.3 Recursive Pose Estimation (SCAAT)

The on-line measurements (section 5.2) are used to estimate the pose of the HiBall in real time, on line. The 1991 system collected a group of similar measurements for a variety of LEDs and sensors, and then used a method of simultaneous non-linear equations called *Collinearity* [4] to estimate the pose of the sensor fixture shown in Figure 1 (left). There was one equation for each measurement, expressing the constraint that a ray from the front principle point of the sensor lens to the LED, must be collinear with a ray from the rear principle point to the intersection with the sensor. Each estimate made use of a group of measurements (typically 20 or so) that together over-constrained the solution.

This *multiple constraint* method had several drawbacks. First, it had a significantly lower estimate rate due to the need to collect multiple measurements per estimate. Second, the system of non-linear equations did not account for the fact that the sensor fixture continued to move throughout the collection of the sequence of measurements. Instead the method effectively assumes that the measurements were taken simultaneously. The violation of this *simultaneity assumption* could introduce significant error during even moderate motion. Finally, the method provided no means to identify or handle unusually noisy individual measurements. Thus, a single erroneous measurement could cause an estimate to jump away from an otherwise smooth track.

In contrast, the approach we use with the new HiBall system produces tracker reports as each new measurement is made rather than waiting to form a complete collection of observations. Because single measurements under-constrain the mathematical solution, we refer to the approach as Single-Constraint-at-a-Time or SCAAT tracking [28, 29]. The key is that the single measurements provide *some* information about the user's state, and thus can be used to incrementally improve a previous estimate. Using a Kalman filter [15] we intentionally fuse measurements that do not individually provide sufficient information, incorporating each individual measurement immediately as it is obtained. With this approach we are able to generate estimates more frequently, with less latency, with improved accuracy, and we are able to effectively estimate the LED positions on-line concurrently while tracking the HiBall (section 5.4).

We use a Kalman filter, a minimum variance stochastic estimator, to estimate the HiBall *state* $\bar{x}$, i.e. the position and orientation of the HiBall. We use a Kalman filter in part because the sensor measurement noise and the typical user motion dynamics can be modeled as normally-distributed random processes, but also because we want an efficient on-line method of estimation. A basic introduction to the Kalman filter can be found in Chapter 1 of [17], while a more complete introductory discussion can be found in [20], which also contains some interesting historical narrative. More extensive references can be found in [7, 12, 14, 16, 17, 30]. The Kalman filter has been used previously to address similar or related problems. See for example [2, 3, 9, 10, 18, 23], and most recently [11].

The SCAAT approach in particular is described in great detail in [28, 29]. The benefits of using this approach, as opposed to a multiple-constraint approach such as [4], are also discussed in [28, 29]. However one key benefit warrants discussion here. There is a direct relationship between the complexity of the estimation algorithm, the corresponding execution time per estimation cycle, and the character of HiBall motion between estimation cycles. As the algorithmic complexity increases, the execution time increases, which allows for significant non-linear HiBall motion between estimation cycles, which in turn implies the need for a more complex estimation algorithm.

The SCAAT approach on the other hand is an attempt to reverse this cycle. Because we intentionally use a single constraint per estimate, the algorithmic complexity is drastically reduced, which reduces the execution time, and hence the amount of motion between estimation cycles. Because the amount of motion is limited we are able to use a simple dynamic (process) model in the Kalman filter, which further simplifies the computations. In short, the simplicity of the approach means it can run very fast, which means it can produce estimates very rapidly, with low noise.

The Kalman filter requires both a model of the process dynamics, and a model of the relationship between the process state and the available measurements. In part due to the simplicity of the SCAAT approach we are able to use a very simple process model. We model the continuous change in the HiBall state vector $\bar{x}(t)$ with the simple differential equation

$$\frac{d}{dt}\bar{x}(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \bar{x}_1(t) \\ \bar{x}_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \mu \end{bmatrix} u(t),$$

where the scalar

$$\bar{x}_2(t) = \frac{d}{dt}\bar{x}_1(t),$$

$u(t)$ is a normally-distributed scalar white noise process, and the scalar μ represents the magnitude of the noise (the spectral density). A similar model with a distinct noise magnitude μ is used for each of the six position and orientation elements. The individual noise magnitudes are determined using an off-line simulation of the system and a non-linear optimization strategy that seeks to minimize the variance between the estimate pose and a known motion path. (See section 6.2.2.) The above differential equation represents a continuous integrated random walk, or an integrated *Wiener* or *Brownian-motion* process. Specifically, we model each component of the linear and angular HiBall velocities as random walks, and use these, assuming constant inter-measurement velocity, to estimate the six elements of the HiBall pose at time $t + \delta t$ as follows:

$$\bar{x}(t + \delta t) = \begin{bmatrix} 1 & \delta t \\ 0 & 1 \end{bmatrix} \bar{x}(t). \qquad (1)$$

In addition to a relatively simple process model, the HiBall measurement model is relatively simple. For any Ceiling LED (section 4.1) and HiBall camera view (section 4.2), the 2D sensor measurement can be modeled as

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \bar{c}_x / \bar{c}_z \\ \bar{c}_y / \bar{c}_z \end{bmatrix} \qquad (2)$$

where

$$\begin{bmatrix} \bar{c}_x \\ \bar{c}_y \\ \bar{c}_z \end{bmatrix} = VR^T(\dot{l}_{xyz} - \bar{x}_{xyz}),$$

V is the camera viewing matrix from section 5.1, the vector $\dot{l}$ contains the position of the LED in the world, and R is a rotation matrix constructed from the orientation quaternion contained in the state vector:

$$R = \text{rot\_from\_quat}(\bar{x}_q).$$

In practice we maintain the orientation of the HiBall as a combination of a global (external to the state) quaternion and a set of incremental angles as described in [28, 29].

Because the measurement model is non-linear we use an *extended Kalman filter*, making use of the Jacobian of the non-linear HiBall measurement model to transform the covariance of the Kalman filter. While this approach does not preserve the Gaussian nature of the covariance, it has been used successfully in countless applications since the introduction of the (linear) Kalman filter. Based on observations of the statistics of the HiBall filter residuals, the approach also appears to work well for the HiBall.

At each estimation cycle, the next of the 26 possible views is chosen randomly. Four points corresponding to the corners of the LEPD sensor associated with that view are then projected into the world using the 3 by 4 viewing matrix for that view, along with the current estimates for the HiBall position and orientation. This projection, which is the inverse of the measurement relationship described above, results in four rays extending from the sensor into the world. The intersection of these rays and the approximate plane of the Ceiling determines a 2D bounding box on the Ceiling, within which are the candidate LEDs for the current camera view. One of the candidate LEDs is then chosen in a least-recently-used fashion to ensure a diversity of constraints.

Once a particular view and LED have been chosen in this fashion, the CIB (section 4.3) is instructed to flash the LED and take a measurement as described in section 5.2. This single measurement is compared with a prediction obtained using (2), and the difference or *residual* is used to update the filter state and covariances using the *Kalman gain* matrix. The Kalman gain is computed as a combination of the current filter covariance, the measurement noise variance (section 6.2.1), and the Jacobian of the measurement model.

A more detailed discussion of the HiBall Kalman filter and the SCAAT approach is beyond the scope of this paper. For additional information see [28, 29].

5.4 On-line LED Autocalibration

Along with the benefit of simplicity and speed, the SCAAT approach offers the additional capability of being able to estimate the 3D positions of the LEDs in the world concurrently with the pose of the HiBall, on line, in real time. This capability is a tremendous benefit in terms of the accuracy and noise characteristics of the estimates. Accurate LED position estimates is so important that prior to the introduction of the SCAAT approach a specialized off-line approach was developed to address the problem [13].

The method we use for autocalibration involves effectively defining a distinct Kalman filter for each and every LED. Specifically, for each LED we maintain the state $\hat{l}$ (estimate of the 3D position) and a 3x3 Kalman filter covariance. At the beginning of each estimation cycle we augment the HiBall Kalman filter described in section 5.3 with the appropriate individual LED filter. In particular we add the three elements of $\hat{l}$ to the state $\bar{x}$, and similarly augment the Kalman filter error covariance matrix with that of the LED filter. We then follow the normal steps outlined in section 5.3, with the result being that the LED portion of the filter state and covariance is updated in accordance with the measurement residual. At the end of the cycle we extract the LED portions of the state and covariance from the filter, and save them externally. The effect is that as the system is being used, it continually refines its estimates of the LED positions, thereby continually improving its estimates of the HiBall pose. Again, for additional information see [28, 29].

5.5 Initialization and Re-Acquisition

The recursive nature of the Kalman filter, and hence the method described in section 5.3, typically requires that the filter be initialized with a known state and corresponding covariance before steady-state operation can begin. This is true for the HiBall system, as convergence cannot be assured from a randomly chosen state. Such an initialization must take place prior to any tracking session, but also upon the (rare) occasion when the filter diverges and "loses lock" as a result of blocked sensor views for example.

Acquiring lock is complicated by the fact that each LEPD sees a number of different widely separated views. Therefore detecting a beacon provides at best an ambiguous set of potential beacon directions in HiBall coordinates. Moreover, before establishing lock, no assumptions can be made to limit the search space of visible beacons. As such, a relatively slow brute-force algorithm is used to acquire lock.

We begin with an exhaustive beacon scan of sufficiently fine granularity to ensure that the central primary field of view is not missed. For the present Ceiling, we flash every 13th LED in sequence, and look for it with the central LEPD until we get a hit. Then a sufficiently large patch of beacons, centered on the hit, is sampled to ensure that several of the views of the central LEPD will be hit. The fields of view are then disambiguated by estimating the yaw of the HiBall from the initial hits, and finally, more selective measurements are made to refine the estimate sufficiently to switch into tracking mode.

6. RESULTS

Three days after the individual pieces of hardware were shown to be functioning properly we demonstrated a complete working system. After months of subsequent tuning and optimization, the system continues to perform both qualitatively and quantitatively as well, or in some respects *better*, than we had anticipated (section 6.1). The articulation of this success is not meant to be self-congratulatory, but to give credit to the extensive and careful modeling and simulation performed prior to assembly (section 6.2). In fact, the Kalman filter parameters found by the optimization procedure described in section 6.2.2 were, and continue to be, used directly in the working system. Likewise much of the software written for the original simulations continues to be used in the working system.

6.1 On-Line Operation

The HiBall system is in daily use as a tool for education and research. For example, it was recently used by Martin Usoh et al. to perform Virtual Reality experiments comparing virtual "flying", walking in place, and real walking [22]. The researchers used the HiBall system to demonstrate that as a mode of locomotion, real walking is simpler, more straightforward, and more natural, than both virtual flying and walking in place. Some images from these experiments are shown in color plate image Welch 3. The unprecedented combination of large working volume and high performance of the HiBall system led the researchers to claim that there was literally nowhere else that they could have meaningfully performed the experiments. Other researchers who visit and try the HiBall system almost always ask how they can get one. We are working on a means to make that happen in selected laboratories.

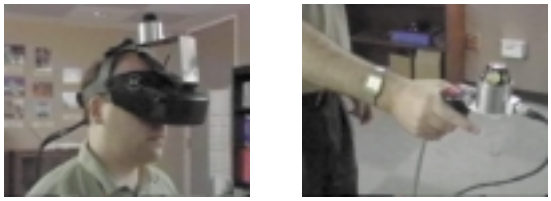


Figure 5. The HiBall being used by Kevin Arthur to track the head and hand for the presence experiments in [22].

Strangely enough, in some sense it is amazing that the system works at all. In particular, you effectively have to know where the HiBall is for it to work at all. The reason is that as a result of a mechanical design trade-off, each sensor field of view is less than six degrees. A small mistake could quickly cause problems. The focal length is set by the size of the sensor housing, which is set by the diameter of the sensors themselves. Energetics enters in also because you have to have light collecting area. And yet the system is amazingly robust: users can dance around, crawl on the floor, lean over, even wave their hands in front of the sensors, and the system does not lose lock. During one session we were using the HiBall as a 3D *digitization probe*, a HiBall on the end of a pencil-shaped fiberglass wand as

shown in Figure 6. We laid the probe down on a table at one point, and were amazed to later notice that it was still tracking, even though it could only "see" 3 or 4 LEDs near the edge of the Ceiling. We picked the probe up and continued using it. It never missed a beat.

The simplest quantitative measurement of tracking system performance is standard deviation of its estimates when it is held stationary. With a tracker as sensitive as the HiBall it is important to be certain that it really is stationary. The raised floor in our laboratory allows motion, for example when a person walks by, that is larger than the expected error in the HiBall. We made careful measurements by resting the support for the HiBall on the concrete sub-floor in our laboratory. The standard deviation of the error on the HiBall estimates while stationary is about 0.2 millimeters and 0.03 degrees. The distribution of the errors fit a normal distribution quite well.

To make measurements of the noise when the HiBall is in motion we rely on the assumption that almost all of the signal resulting from normal human motion is at frequencies below 2 Hz. We use Welch's method [31] to isolate the signal energy above 2 Hz. (Since the SCAAT method is running at about 2000 Hz it is reasonable to assume that most of the noise energy is above 2 Hz.) This measurement produces results that are comparable to those made with the HiBall stationary, except at positions for which there are very few LEDs visible in only one or two views. In these positions, near the edge of the ceiling, the geometry of the constraints results in amplification of errors. For nearly all of the working volume of the tracker the standard-deviation of the noise on measurements while the HiBall is still or moving is 0.2 millimeters and 0.03 degrees.

In July of 1999 two Boeing Corporation engineers, David Himmel and David Princehouse, visited our laboratory to assess the accuracy of the HiBall system in tracking the position and orientation of a hand-held pneumatic drill. The engineers are interested in improving the accuracy of holes drilled by hand during the aircraft manufacturing process. To asses the accuracy of the HiBall system they brought with them an aluminum "coupon" (plate) with holes pre-drilled at locations accurate to 1/1000 of an inch. In a set of carefully controlled experiments we together measured an average positioning error of 1/2 millimeter (20/1000 inch) at the tip of the HiBall probe mentioned above and shown in the left image of Figure 6. Unfortunately at the time we decided *not* to position the experimental platform on the concrete sub-floor, but for expediency, to live with the additional uncertainty of placing it on the raised floor. However we are encouraged by the results, and are excited about the possibility that the HiBall system has uses beyond Virtual Reality tracking. Some images from the Boeing experiments are shown in Figure 6 and in color plate image Welch 3.

Figure 6. Left: a Boeing engineer uses our 3D *digitization probe* (HiBall on a pencil-shaped fiberglass rod) to measure the pre-drilled holes in their aluminum coupon. Right: the HiBall is mounted on a hand-held pneumatic drill for additional testing and measurements.

6.2 Off-Line Simulation and Modeling

The design of the HiBall system made substantial use of simulation, in some domains to a very detailed level of abstraction. For example, Zemax [32] was used extensively in the design and optimization of the optical design, including the design of the filter glass lenses, and geometry of the optical component layout. AutoCAD™ was used to design, specify, and fit-check the HiBall body mechanicals, to visualize the physical design, and to transmit the design to our collaborators at the University of Utah for fabrication by the *Alpha 1* System [1, 21]. A custom ray-tracing system was built by Stefan Gottschalk (UNC) for the purpose of evaluating the optical behavior and energetics of the primary, secondary, and tertiary fields of view; the results were used by the noise model developed by Chi [8] as described in the next section.

In addition, a complete simulator of the system was written in C++. This simulator, discussed further in section 6.2.2, was used to evaluate the speed, accuracy, and robustness of the system. In addition it was used to "tune" the Kalman filter for realistic motion dynamics. This simulator continues to be used to evaluate mechanical, optical, and algorithmic alternatives.

6.2.1 HiBall Measurement Noise Model

Signal-to-noise performance is a prime determiner of both accuracy and speed of the system, so an in-depth study [8] was performed to develop a detailed noise model accounting for properties of the LED, the LEPD (sensor), the optical system, the physical distance and pose, the electronics, and the dark-light-dark integrations described in section 5.2. The predominant noise source is shot noise, with Johnson noise in the sheet resistivity of the LEPD surfaces being the next most significant. Careful measurements made in the laboratory with the actual devices yielded results that were almost identical to those predicted by the sophisticated model in [8]. A simplified version of this model is used in the real system to predict the measurement noise for the Kalman filter (section 5.3) when the automatic gain control described in section 5.2 is not in use.

6.2.2 Complete System Simulations

To produce realistic data for developing and tuning our algorithms we collected position and orientation reports from our first generation ceiling tracker at its 70 Hz maximum report rate. These data were recorded both from

naive users visiting our monthly "demo days" and from experienced users in our labs. Based on our previous research [5] we filtered the raw ceiling data with a non-causal zero-phase-shift low-pass filter to eliminate energy above 2 Hz. The output of the low-pass filtering was then re-sampled at whatever rate we wanted to run the simulated tracker, usually 1000 Hz. For the purposes of our simulations these paths were considered to be a perfect representation of the user's motion. Tracking error was determined by comparing this input "truth" to the estimate produced by the tracker.

The simulator reads camera models describing the 26 views, the sensor noise parameters, the LED positions and their expected error, and the motion path described above. Before beginning the simulation, the LED positions are perturbed from their ideal positions by adding normally distributed error to each axis. Then, for each simulated cycle of operation, the "true" position and orientation are updated using the input motion path. Next, a view is chosen and a visible LED within that view is selected, and the image-plane coordinates of the LED on the chosen sensor are computed using the camera model for the view and the LED as described in section 5.3. These sensor coordinates are then perturbed based on the sensor noise model (section 6.2.1) using the distance and angle to the LED. Now these noise corrupted sensor readings are fed to the SCAAT filter to produce an updated position estimate. The position estimate is compared to the true position to produce a scalar error metric described next.

The error metric we used combines the error in position and orientation in a way that relates to the effects of tracker error on a HMD user. Imagine points arrayed around the user at some fixed distance. We compute two sets of coordinates for these points; the true position using the true pose, and their estimated position using the estimated pose. The error metric is then the sum of the distances between the true and estimated positions of these points. By adjusting the distance of the points from the user we can control the relative importance of the orientation and the position error in the combined error metric. If the distance is small, then the position error is weighted most heavily; if the distance is large then the orientation error is weighted most heavily. Our single error metric for the entire run is the square-root of the sum of the squares of all the distances.

Determining the magnitude of the parameters which control the SCAAT Kalman filter is called tuning. We used Powell's method [19] to minimize the above error metric. Starting with a set of parameters we ran the simulator over a full motion run to determine the total error for the run. Then the optimizer made a small adjustment to the parameters and the process was repeated. These runs required hours of computer time and some skill (and luck) in choosing the initial parameters. Of course, it was important to choose motion paths that were representative of expected user motion. For example, a run in which the user is very still would result in very different tuning from a run in which the user moves very vigorously.

7. FUTURE WORK

At the moment we are investigating the use of a multi-modal or *multiple-model* Kalman filter framework [6, 7]. The reason is because the current filter form (section 5.3) and tuning values (section 6.2.2) are a compromise between the responsiveness desired for high dynamics, and the heavy filtering desired for smooth estimates during very slow or no motion. As it stands, the system is *almost* smooth enough for open-loop Augmented Reality applications such as computer-assisted medical procedures. The problem is that when a user attempts to sit very still, for example to align a needle with a tumor for a biopsy, even the small noise of our system results in visually-noticeable jitter. A multiple-model implementation of the HiBall should be able automatically, continuously, and smoothly choose between one Kalman filter tuned for high dynamics, and another tuned for little or no motion. We have this working in simulation, but not yet implemented.

As mentioned in section 4.3, the system was designed to support wireless communication between the HiBall and the CIB, without significant modification or added information overhead. However because commercial head-mounted displays are themselves tethered at this time, we have felt little incentive to actually implement a wireless version of the system. As it turns out, our users are becoming increasingly frustrated by the cumbersome cabling that you must drag with you when walking around our laboratory. As such we are now beginning work on a completely wireless HiBall and display system.

Beyond improving the existing system, we are anxious to head down a path of research and development that will lead to systems with reduced dependency on the laboratory infrastructure. For example, our current Ceiling panel design with 32 LEDs per panel, provides far more dense coverage than we believe is necessary. The density of LEDs is a result of design based on the original sensor fixture show in Figure 1, and the original multiple-constraint algorithm [4]. Furthermore, we believe that we could achieve similar performance using a version of the HiBall that has a small number of *wide field of view* cameras.

While we are very happy with our ability to autocalibrate the LED positions concurrently while tracking (section 5.4), we would eventually like to take this one step further and begin with *no* estimates of beacon locations, and possibly no notion of individual beacon identity. We have done some preliminary investigation and simulation that indicates this should be possible. Such capability could drastically reduce the cost of the system, and provide immense flexibility.

Finally, we believe that by leveraging the knowledge gained from successful work in the laboratory, it may some day be possible to achieve similar performance with no explicit infrastructure, anywhere in a building, or even outdoors. A particular approach that we are interested in pursuing is the hybrid approach presented in [27]. It is our belief that no single technology or fundamental modality can provide the necessary performance, in a consistent and robust fashion,

across the necessary range of dynamics for head-mounted displays. We believe that the solution must lie in the clever and careful combination of complementary technologies.

8. ACKNOWLEDGEMENTS

We want to acknowledge several current Tracker Project members, including (alphabetically) Matthew Cutts, Henry Fuchs, Jeffrey Juliano, Benjamin Lok, John Thomas, Nicholas Vallidis, Hans Weber, and Mary Whitton. We also want to thank several past members and contributors. In particular we want to acknowledge the many contributions of past members Philip Winston, Scott Williams, and Pawan Kumar. We also want to acknowledge the many early contributions of previous members Ronald Azuma, Stefan Gottschalk, and Jih-Fang Wang [25], and the contributions of Al Barr (California Institute of Technology) and John "Spike" Hughes (Brown University) to the original off-line LED calibration work [13] that led to the simpler Ceiling panels shown in Figure 3. Finally we want to acknowledge our many collaborators in the NSF Science and Technology Center for Computer Graphics and Scientific Visualization, specifically our collaborators in mechanical design and fabrication at the University of Utah: Rich Riesenfeld, Sam Drake, and Russ Fish.

This work was supported in part by DARPA/ETO contract number DABT 63-93-C-0048, "Enabling Technologies and Application Demonstrations for Synthetic Environments", Principle Investigators Frederick P. Brooks Jr. and Henry Fuchs (University of North Carolina at Chapel Hill), and by the National Science Foundation Cooperative Agreement no. ASC-8920219: "Science and Technology Center for Computer Graphics and Scientific Visualization," Center Director Andy van Dam (Brown University). Principle Investigators Andy van Dam, Al Barr (California Institute of Technology), Don Greenberg (Cornell University), Henry Fuchs (University of North Carolina at Chapel Hill), Rich Riesenfeld (University of Utah).

9. REFERENCES

[1] "Alpha 1 Publications," University of Utah, Department of Computer Science. [Cited May 28, 1999]. Available from http://www.cs.utah.edu/projects/alpha1/a1_publications.html.

[2] Azarbayejani, Ali, and Alex Pentland. June 1995. "Recursive Estimation of Motion, Structure, and Focal Length," *IEEE Trans. Pattern Analysis and Machine Intelligence*, June 1995, 17(6).

[3] Azuma, Ronald. 1995. "Predictive Tracking for Augmented Reality," Ph.D. dissertation, University of North Carolina at Chapel Hill, TR95-007.

[4] Azuma, Ronald and Mark Ward. 1991. "Space-Resection by Collinearity: Mathematics Behind the Optical Ceiling Head-Tracker," UNC Chapel Hill Department of Computer Science technical report TR 91-048 (November 1991).

[5] Azuma, Ronald and Gary Bishop. 1995. "A Frequency-Domain Analysis of Head-Motion Prediction," Proceedings of SIGGRAPH 95 (Los Angeles, CA), August 6-11 1995. pp. 401-408.

[6] Bar-Shalom, Yaakov, and Xiao-Rong Li. 1993. *Estimation and Tracking: Principles, Techniques, and Software*. Artec House, Inc.

[7] Brown, R.G. and P. Y. C. Hwang. 1992. *Introduction to Random Signals and Applied Kalman Filtering*, 2nd Edition, John Wiley & Sons, Inc.

[8] Chi, Vernon L. "Noise Model and Performance Analysis Of Outward-looking Optical Trackers Using Lateral Effect Photo Diodes," TR95-012, Department of Computer Science, UNC at Chapel Hill, April 1995.

[9] Emura, S. and S. Tachi. 1994. "Sensor Fusion based Measurement of Human Head Motion," *Proceedings 3rd IEEE International Workshop on Robot and Human Communication, RO-MAN'94 NAGOYA* (Nagoya University, Nagoya, Japan).

[10] Foxlin, Eric. 1993. "Inertial Head Tracking," Master's Thesis, Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

[11] Foxlin, Eric, Michael Harrington, George Pfeifer. 1998. "Constellation™: A Wide-Range Wireless Motion-Tracking System for Augmented Reality and Virtual Set Applications," SIGGRAPH 98 Conference Proceedings, Annual Conference Series. ACM SIGGRAPH, July 19-24, 1998, Orlando, FL.

[12] Gelb, A. 1974. *Applied Optimal Estimation*, MIT Press, Cambridge, MA.

[13] Gottschalk, Stefan, and John F. Hughes. 1993. "Autocalibration for Virtual Environments Tracking Hardware," Proceedings of Annual Conference Series, ACM SIGGRAPH 93 (Anaheim, CA, 1993).

[14] Jacobs, O. L. R. 1993. *Introduction to Control Theory*, 2nd Edition. Oxford University Press.

[15] Kalman, R.E. 1960. "A New Approach to Linear Filtering and Prediction Problems," Transaction of the ASME—Journal of Basic Engineering, pp. 35-45 (March 1960).

[16] Lewis, Richard. 1986. *Optimal Estimation with an Introduction to Stochastic Control Theory*, John Wiley & Sons, Inc.

[17] Maybeck, Peter S. 1979. *Stochastic Models, Estimation, and Control*, Volume 1, Academic Press, Inc.

[18] Mazuryk, Thomas and Michael Gervautz. 1995. "Two-Step Prediction and Image Deflection for Exact Head Tracking in Virtual Environments," *EUROGRAPHICS '95*, Vol. 14, No. 3, pp. 30-41.

[19] Press, William H., Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. 1990. *Numerical Recipes in C, The Art of Scientific Computing*, Cambridge University Press.

[20] Sorenson, H.W. 1970. “Least-Squares estimation: from Gauss to Kalman,” *IEEE Spectrum*, Vol. 7, pp. 63-68, July 1970.

[21] Thomas, S. W., “The Alpha_1 Computer-Aided Geometric Design System in the Unix Environment,” in Proc. Computer Graphics and Unix Workshop (Katz, L., ed.), USENIX Organization, Dec. 1984.

[22] Usoh, Martin, Kevin Arthur, Mary C. Whitton, Rui Bastos, Anthony Steed, Mel Slater, Frederick P. Brooks, Jr. “Walking > Walking-in-Place > Flying, in Virtual Environments,” to appear in SIGGRAPH 99 Conference Proceedings, Annual Conference Series. ACM SIGGRAPH, August 1999, Los Angeles, CA.

[23] Van Pabst, J. V. L. and Paul F. C. Krekel. “Multi Sensor Data Fusion of Points, Line Segments and Surface Segments in 3D Space,” TNO Physics and Electronics Laboratory, The Hague, The Netherlands. [cited 19 November 1995]. Available from http://www.bart.nl/~lawick/index.html.

[24] Wang, Jih-fang, Ronald Azuma, Gary Bishop, Vernon Chi, John Eyles, Henry Fuchs, “Tracking a Head-mounted Display in a Room-sized environment with Head-mounted Cameras,” Proceedings of the SPIE 1990 Technical Symposium on Optical Engineering and Photonics in Aerospace Sensing, Orlando, Florida, 16-20 April 1990.

[25] Wang, Jih-fang, Vernon Chi, Henry Fuchs, “A Real-time Optical 3D Tracker for Head-mounted Display Systems,” Proceedings of the 1990 Symposium on Interactive 3D Graphics, Snowbird, UT, 25-28 March 1990; in Computer Graphics, Vol. 24, No. 2, March, 1990, pp. 205-215.

[26] Ward, Mark, Ronald Azuma, Robert Bennett, Stefan Gottschalk, and Henry Fuchs. 1992. “A Demonstrated Optical Tracker With Scalable Work Area for Head-Mounted Display Systems,” *Proceedings of 1992 Symposium on Interactive 3D Graphics* (Cambridge, MA, 29 March - 1 April 1992), pp. 43-52.

[27] Welch, Gregory. 1995. “Hybrid Self-Tracker: An Inertial/Optical Hybrid Three-Dimensional Tracking System,” University of North Carolina, Department of Computer Science, TR 95-048.

[28] Welch, Gregory and Gary Bishop. 1997. “SCAAT: Incremental Tracking with Incomplete Information,” SIGGRAPH 97 Conference Proceedings, Annual Conference Series. ACM SIGGRAPH, August 1997, Los Angeles, CA.

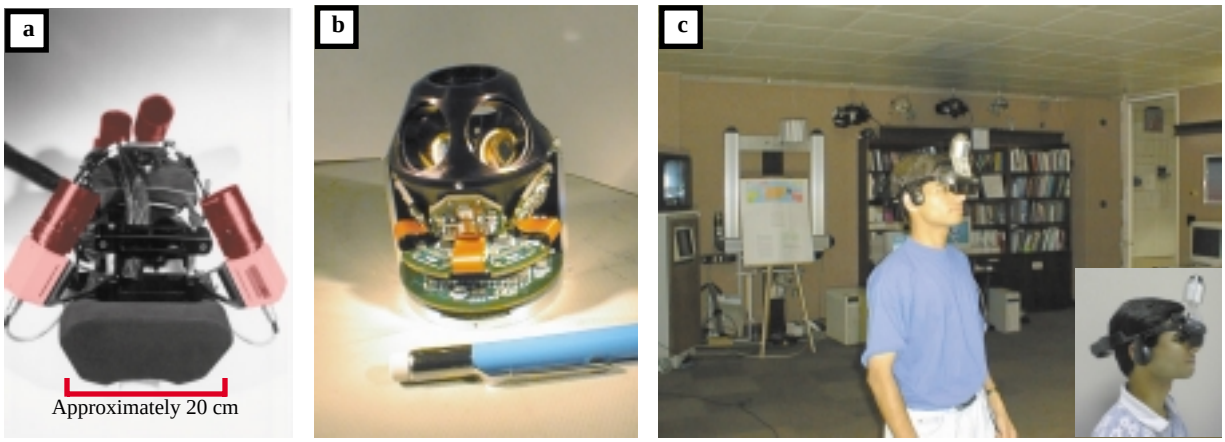
[29] Welch, Gregory. 1996. “SCAAT: Incremental Tracking with Incomplete Information,” University of North Carolina at Chapel Hill, doctoral dissertation, Technical Report TR 96-051.

[30] Welch, Gregory and Gary Bishop. 1995. “An Introduction to the Kalman Filter,” University of North Carolina, Department of Computer Science, TR 95-041.

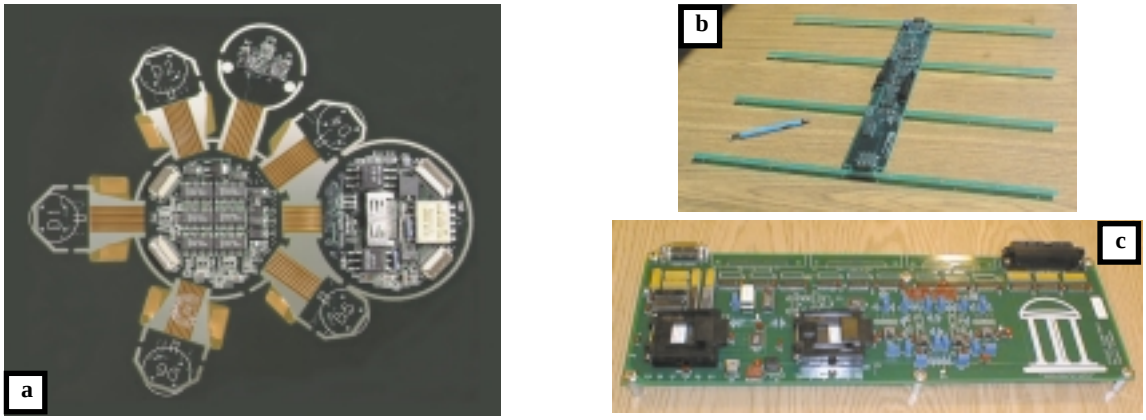
[31] Welch, P.D. “The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms.” IEEE Trans. Audio Electroacoust. Vol. AU-15 (June 1967). pp. 70-73.

[32] ZEMAX Optical Design Program User's Guide, Version 4.5, Focus Software, Inc., Tucson, AZ, 1995.

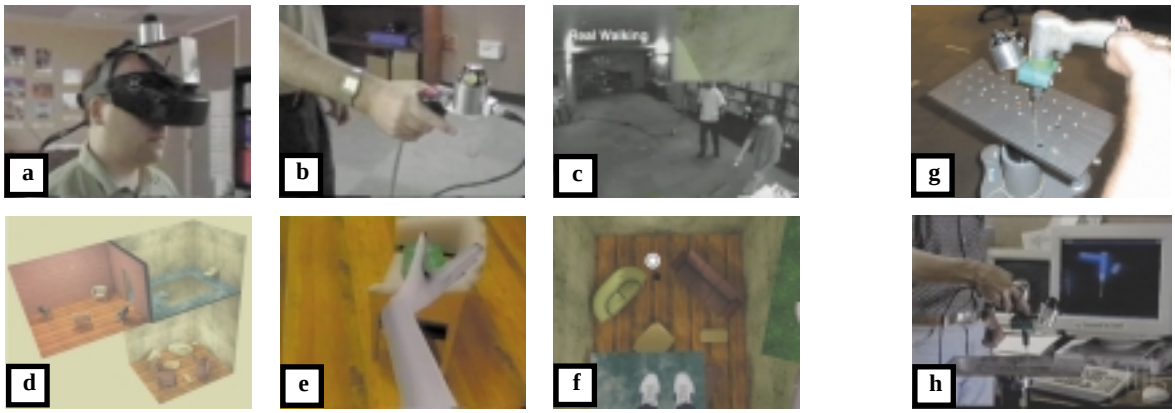
COLOR PLATE



Welch 1: The left image (a) shows the head-mounted display and bulky sensor fixture from the original system. The sensors are colorized in red. The middle image (b) shows a close-up of the new self-contained HiBall, with lenses and part of the cover removed. The right image (c) shows graduate student Pawan Kumar using the new HiBall system.



Welch 2: Image (a) shows an unassembled HiBall printed circuit board. The board has two separate large round sections for analog and digital circuits, and six small round sections for sensors. These rigid PCB sections are connected by flexible trace segments to eliminate connectors. After populating the board, the pieces are cut away from the rectangular board and folded into the HiBall as shown. Image (b) shows the Ceiling panel PC board from Figure 3, and (c) the CIB from Figure 4.



Welch 3: The images (a)-(f) show the HiBall system in use by Usoh and Martin and their colleagues for experiments in [22]. Images (a)-(b) show the HiBall being used to track the head and hand; (c) shows a test subject walking under the HiBall Ceiling; (d) is an overview of the virtual world; (e)-(f) are sample test subject views. Image (g) shows the HiBall mounted on a hand-held pneumatic drill for experiments with Boeing engineers; (h) shows a HiBall-based drill-guiding application.