HISTORY: The Use of the Kalman Filter for Human Motion Tracking in Virtual Reality

Abstract

In 1960 Rudolph E. Kalman published his now famous article describing a recursive solution to the discrete-data linear filtering problem (Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME—Journal of Basic Engineering, 82* (D), 35–45, 1960). Since that time, due in large part to advances in digital computing, the *Kalman filter* has been the subject of extensive research and applications, particularly in the area of autonomous or assisted navigation. The purpose of this paper is to acknowledge the approaching 50th anniversary of the Kalman filter with a look back at the use of the filter for human motion tracking in virtual reality (VR) and augmented reality (AR).

In recent years there has been an explosion in the use of the Kalman filter in VR/AR. In fact, at technical conferences related to VR these days, it would be unusual to see a paper on tracking that did not use some form of a Kalman filter, or draw comparisons to those that do. As such, rather than attempt a comprehensive survey of all uses of the Kalman filter to date, what follows focuses primarily on the early discovery and subsequent period of evolution of the Kalman filter in VR, along with a few examples of modern commercial systems that use the Kalman filter.

This paper begins with a very brief introduction to the Kalman filter, a brief look at the origins of VR, a little about *tracking* in VR—in particular the work and conditions that gave rise to the use of the filter, and then the evolution of the use of the filter in VR.

1 The Kalman Filter

The Kalman filter is a set of mathematical equations that provides an efficient computational means to recursively estimate the state and error covariance of a process, in a way that minimizes the mean of the squared error covariance. Specifically, the Kalman filter addresses the general problem of trying to estimate the state $\bar{x} \in \Re^n$ and error covariance $P \in \Re^{n \times n}$ of a discrete-time controlled process that is governed by the linear stochastic difference equation

$$\bar{x}_k = A\bar{x}_{k-1} + \bar{w}_k , \qquad (1)$$

and observable via measurements $\bar{x} \in \Re^m$ modeled by

$$\bar{z}_k = H\bar{x}_k + \bar{v}_k , \qquad (2)$$

where $A \in \Re^{n \times n}$ models the transition of the state over time, $H \in \Re^{m \times n}$ models the relationship between the state and the measurements, and $\bar{w}_k \in \Re^n$ and $\bar{v}_k \in \Re^m$ model zero-mean, normally distributed, and spectrally white process and measurement noise vectors.

The recursive operation of the filter involves the use of a repeating cycle of prediction and correction using Equations 1 and 2. Specifically, when a measurement $\bar{z}_k$ becomes available at discrete time step k, one predicts the a priori state $\tilde{x}_k^-$ and error covariance P_k^- using

$$\tilde{x}_k^- = A\bar{x}_{k-1} \qquad (3)$$

$$P_k^- = AP_{k-1}A^T + Q , \qquad (4)$$

where Q is the expected covariance of $\bar{w}_k$, then computes the gain matrix $K \in \Re^{n \times m}$ using

$$K = P_k^- H^T (HP_k^- H^T + R)^{-1} , \qquad (5)$$

Gregory F. Welch
The University of North Carolina at Chapel Hill
Department of Computer Science
CB 3175, Sitterson Hall
Chapel Hill, NC 27510-3175
welch@cs.unc.edu

and then corrects the state and error covariance to obtain the a posteriori versions using

$$\tilde{x}_k = \tilde{x}_k^- + K(\bar{z}_k - H\tilde{x}_k^-) \qquad (6)$$

$$P_k = (I - KH)P_k^- \qquad (7)$$

The Kalman filter is very powerful in several respects: it supports estimations of past, present, and future states $\bar{x}$, and it can do so even when the state elements are hidden (not directly observable in $\bar{z}$), or the precise nature of the modeled system is unknown. It is optimal in the sense that the gain K minimizes the trace of the error covariance P_k, when the process, measurement, and noise models are appropriate.

A more complete explanation of the Kalman filter is beyond the scope of this paper. The reader is encouraged to consult one of the many books or papers describing the filter, such as Gelb (1974); Welch and Bishop (1995); Brown and Hwang (1996).

2 Virtual Reality

In 1963 Ivan Sutherland introduced Sketchpad, a computer program that used an *x-y* vector display and a tracked light pen for computer-aided drawing. Sketchpad is arguably the first interactive graphical user interface to a computer. Two years later, he described the "ultimate display" for computers as "a room within which the computer can control the existence of matter" (Sutherland, 1965, p. 508). He wrote that "[a] chair displayed in such a room would be good enough to sit in. Handcuffs displayed in such a room would be confining, and a bullet displayed in such a room would be fatal." Subsequently Sutherland and his student, Bob Sproull, created what is usually considered to be the first head-mounted display (HMD) system for interactive computer graphics. Their system generated binocular imagery that was rendered appropriately for the position and orientation of the moving head. As shown in Figure 1, the display was suspended from a counterbalanced telescoping and pivoting mechanical arm, which, with the help of ultrasonic transducers, was used to track the dynamic head pose. This system offered the first example of using computers to allow people to see into another virtual world—what we now call a *virtual reality* (VR).

Today, research and commercial VR systems are used for simulation and training, industrial design, phobia therapy or other health-related applications, surgical planning and assistance, artistic applications, and of course games. Turnkey systems are available that include visuals, sound, haptic and force feedback, including even taser-like electric shocks to simulate being shot.

3 Human Motion Tracking for Virtual Reality

As VR is inherently an interactive paradigm, an essential part of most VR systems is the online, real-time estimation of human motion—primarily head motion—for the purpose of generating head-coupled synthetic imagery. For most applications, head tracking is considered more important than stereo imagery. To quote Sutherland,

> Although stereo presentation is important to the three-dimensional illusion, it is less important than the change which takes place in the image when the observer moves his head. The image presented by the three-dimensional display must change in exactly the same way which the image of a real object would change for similar motions of the user's head. Psychologists have long known that moving perspective images appear strikingly three-dimensional even without stereo presentation. (Sutherland, 1968, p. 757)

Presenting computer-generated imagery that changes "exactly the same way that the image of a real object would change" translates into very demanding requirements for the accuracy, resolution, and speed (low latency and high update rate) of the head tracking and image generation. The problem is that we humans have a lifetime of experience with visual, vestibular, and proprioceptive correspondences, and even the slightest perceptible deviations can result in a break in the user's sense of presence in the virtual world, or worse, sickness.
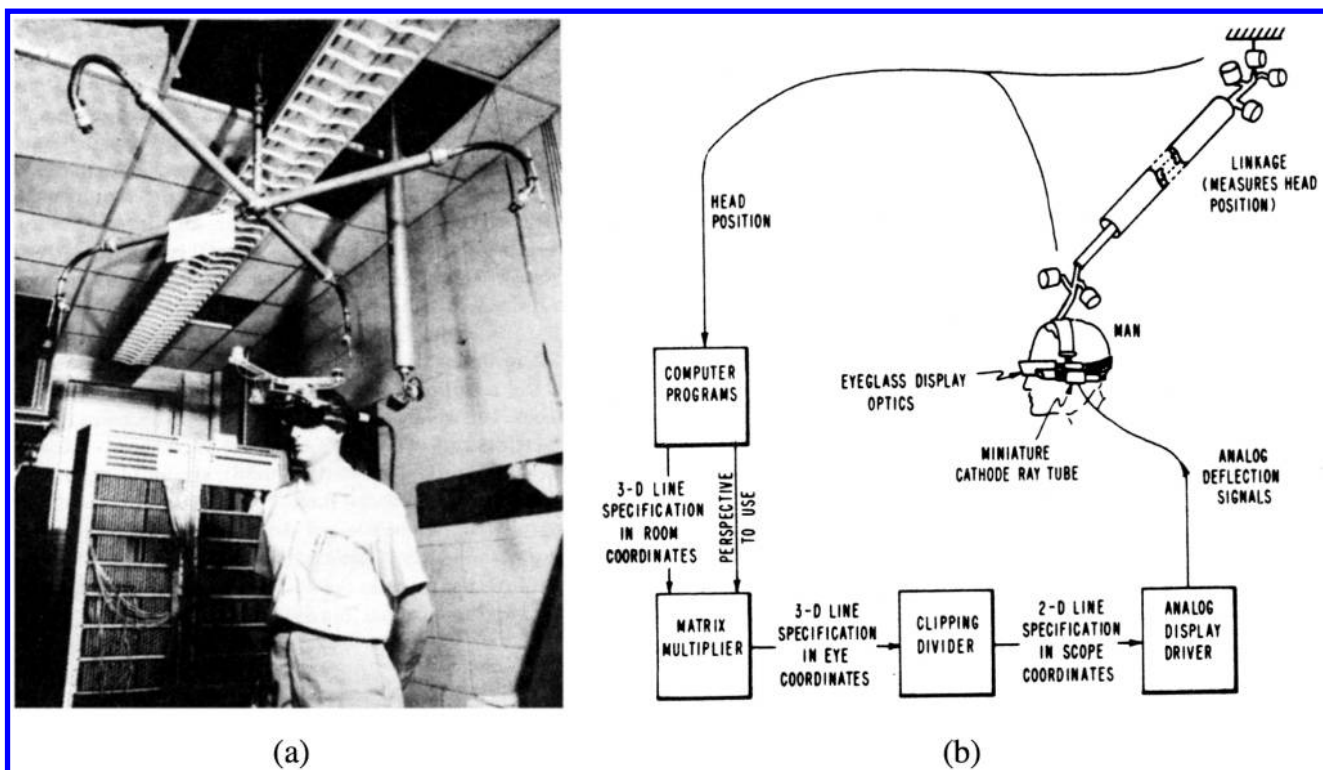
Figure 1. *Ivan Sutherland's head-mounted three-dimensional display (c. 1968). The display was suspended from a counterbalanced telescoping and pivoting mechanical arm, which, with the help of ultrasonic transducers, was used to track the dynamic head pose. (a) The mechanical head-position sensing apparatus in use. (b) The various parts of the three-dimensional display system. This system offered the first example of using computers to allow people to see into another virtual world, what we now call a virtual reality. Images reproduced from Sutherland (1968), with permission of the Charles Babbage Institute.*

The challenges are even greater for *augmented reality* (AR), where computer graphics images are blended with a user's view of the real world. Misregistration can be confusing or distracting, or worse, for applications such as AR-assisted surgery. Sutherland once stated their goal was a resolution of 1/100 of an inch and one part in 10,000 of rotation. Today's systems typically employ user-worn components the size of a golf ball or smaller, and can achieve head position accuracy and resolution of tenths of millimeters, and orientation accuracy and resolution of hundredths of degrees, all with latencies on the order of milliseconds.

While the causes of visual error in interactive computer graphics systems are numerous, the errors originating in the tracking system tend to dominate all other sources. In Rich Holloway's 1995 Ph.D. dissertation, he thoroughly analyzed the sources of error in an aug-

mented reality (AR) system for computer-aided surgical planning (Holloway, 1995). He concluded that error in head tracking is the major cause of registration error in AR systems, and that latency is one of the primary causes of error in head tracking. His analysis is still appropriate today, and will be for the foreseeable future, given the fundamental nature of it.

The sources of error in a tracking system are usually characterized as static or dynamic. With respect to the former, there are typically several components between which static geometric transforms must be estimated prior to use of the system. Examples would be the six-degree-of-freedom (6 DOF) transform from a laboratory coordinate system to that of the tracking system, and the 6 DOF transform from a head-worn tracker sensor to the user's eyes. Depending on the distances and angles involved, the error magnification can be nontrivial.

In addition, some trackers exhibit systematic, repeatable distortions of their measurement volume. Usually these distortions can be modeled and corrected, as long as they remain unchanged between this calibration procedure and run time. Finally, as with any control system, random measurement noise or poor signal can lead to jitter in the pose estimates. The amount of jitter is often proportional to the distance between the sensor(s) and the source(s); however, it can also increase under conditions of poor observability, such as might be caused by sensor occlusions or operation near the edge of a tracker's working volume.

In terms of dynamic error, the dominant source of error is latency combined with continued user motion after a tracker cycle (sample, estimate, produce) has started. This dominance primarily results from the interactive nature of VR. In his 1999 article "What's real about virtual reality?" Fred Brooks stated that "[i]n my opinion, end-to-end system latency is still the most serious technical shortcoming of today's VR systems" (Brooks, 1999, p. 18). Consider that over brief intervals humans can translate heads and hands at several meters per second, and rotate their heads hundreds of degrees per second.

Even a relatively modest head rotation rate of 100 degrees per second corresponds to one tenth of a degree per millisecond. Given that the practical minimum overall VR system latency ranges from 100–200 ms these days, the latency corresponds to 10°–20° of dynamic error, which can produce a perception of "swimming" of the imagery. However, latency can be problematic even for delicate motion. For an arm's length application using an HMD for surgical planning, Holloway arrived at a rule of thumb of 1 ms of latency corresponds to 1 mm of misregistration (Holloway, 1995).

Researchers continue to study the impact of system latency on VR systems; however, given the impossibility of reducing the latency to zero, motion prediction schemes abound. In fact, as the reader will see in the next section, rather than filtering or fusion, it is the problem of motion prediction that gave rise to the earliest uses of the Kalman filter in human motion tracking for virtual reality. Quite a few years passed before VR researchers realized that the Kalman filter is perhaps *the*

perfect tool for elegantly combining multisensor fusion, filtering, and motion prediction in a single fast and accurate framework.

4 The Use of the Kalman Filter in VR

Work on head-tracking systems in general has been going on at least since Sutherland's work in the early 1960s. Research and commercial systems have explored and employed virtually all available physical media: mechanical linkages; light (including passive and active optics, photodiodes, and image-forming cameras); sound (time-of-flight, phase, spread spectrum); magnetics (passive and electromagnetic); inertia (accelerometers and gyros); and radio frequencies (WiFi and global positioning). There exist several nice reviews of the general approaches, including Meyer, Applewhite, and Biocca (1992); Bhatnagar (1993) and more recently, Foxlin (2002b); Welch and Foxlin (2002). What follows is an attempt to tell the story of how the use of the Kalman filter evolved in tracking systems for VR.

As with many things in history, it is difficult to pinpoint a single defining moment when the Kalman filter was first used in VR. Publication dates do not necessarily reflect the exact timing of the work, much less the ideas. However, through careful cataloging of publications, and some personal recollections of other researchers, a chronology has evolved, with some interesting epochs.

4.1 Early Related Work

Part of the difficulty in identifying a first use of the Kalman filter in virtual reality (VR) is that the notion of what VR is, and when it began, is itself a bit subjective. Even if it were possible to identify a precise moment or event, a historical treatment which did not look at the conditions that gave rise to the event, and the related work preceding it, would be incomplete. This section is an attempt to do that.

While Ivan Sutherland is credited with the birth of VR as a general paradigm in 1965, flight simulators are usually considered VR systems (Brooks, 1999), and

their development traces back to the early 1960s if considering digital displays; to the 1950s if considering systems that flew cameras over terrain models; or even to 1929 when Edwin Link developed the first pilot trainer (Wikipedia, 2008). Looking more narrowly at head-slaved systems with interactive graphics, around 1966 the Air Force Aerospace Medical Research Laboratory (AFAMRL) began work on the Visually Coupled Airborne System Simulator (VCASS), one of at least two visually coupled systems facilities being pursued by AFAMRL at that time for in-flight and simulated control of threat and weapons systems. While it is not clear they were using a Kalman filter until the mid-1980s (Haas, 1983, 1984), this work is relevant in that it gave rise to the development of the Space Synchro (SPASYN) magnetic head-tracking technique (DeRuyck & Kuipers, 1973), which would eventually become the primary means for head tracking in VR for years to come.

Figure 2 comes from what appears to be the first publicly available publication describing the VCASS system (Kocian, 1977). The figure shows the magnetic transmitter and helmet-mounted receiver, along with the helmet-mounted display. The SPASYN approach was developed by Polhemus Navigation Services in the early 1970s, specifically for the U.S. Air Force for head tracking in cockpits and cockpit simulators (DeRuyck & Kuipers, 1973). The system operated by sequentially exciting a tri-axial coil transmitter with AC signals while measuring the resulting currents in a user-worn tri-axial coil sensor, providing a 5 DOF estimate of the pilot's head pose (2D head position and 3D head orientation).

Some time around 1979, Polhemus introduced a commercial version called the Isotrak that used magnetic field strength to estimate distance to achieve full 6 DOF pose estimates (Raab, Blood, Steiner, & Jones, 1979). The availability of the Isotrak system marked a turning point for VR researchers in that not only could they actually purchase a turnkey tracking system, but the small, lightweight electromagnetic sensor offered relatively accurate position and orientation. The Isotrak's 1993 successor, the Polhemus Fastrak, is still in use today. An interesting bifurcation occurred in 1986, when Jack Sculley and Ernie Blood (a co-author of
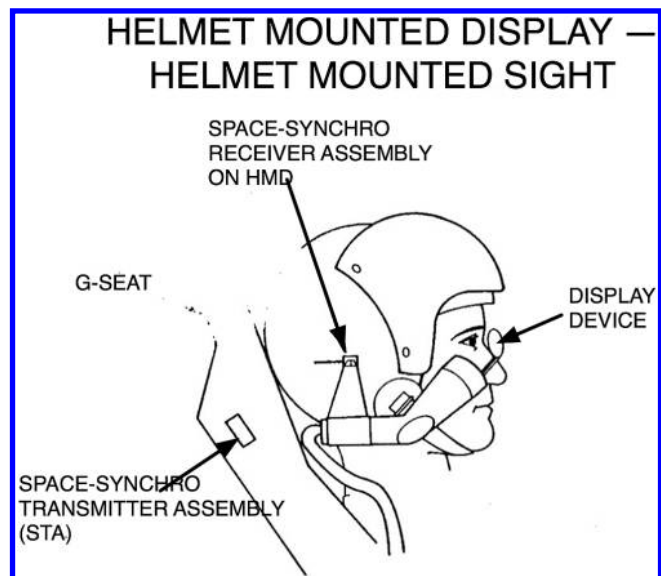
Figure 2. *A reproduction of Figure 2 from Kocian (1977), showing an early head-mounted display developed for the U.S. Air Force's Visually Coupled Airborne Systems Simulator. New figure labels were overlaid to improve on the scanned original. Note the indications of the SPACE-SYNCHRO magnetic receiver and transmitter used to track the position and orientation of the pilot's head. Note also the display device used to place virtual imagery in front of the pilot's right eye. Image reproduced from Kocian (1977).*

Raab et al., 1979) left Polhemus and formed Ascension Technology Corporation. Ascension went on to develop DC magnetic field systems to address interference faced by the AC approach in the presence of ferromagnetic materials. In 1992 Ascension introduced the 6 DOF Bird system, and similar to the Fastrak, the Bird is still in use today.

While the Polhemus Isotrak was groundbreaking in several ways, it turns out that the device latencies were large enough to be raised as a significant issue for numerous users. For example, when Randy Pausch was using it in the late 1980s for his renowned "Virtual reality on five dollars a day" project, he noted, "[t]he major limitation of our system's usability is the lag of the Polhemus Isotrak. Other researchers using the Isotrak have also reported this problem; no one has precisely documented its duration, but it is within 150 and 250 ms." (Pausch, 1991, p. 267). When you add the delays in the subse-

quent synchronization and rendering in the graphics pipeline, the delays were likely on the order of 1/2 s.

In fact, no matter what the source of tracking information, researchers were discovering that the overall VR system latencies were an issue that had to be addressed. Perhaps not surprisingly, such efforts began in the flight simulator community, as a pilot's head could potentially undergo significant head motion dynamics. In the early 1980s, researchers at Williams Air Force Base (U.S. Air Force Human Resources Laboratory, Operations Training Division) were studying the perceptions of overall system latency for an HMD tracked with mechanical linkages and potentiometers. In 1983, Uwe List reported that pilots can rotate their heads with accelerations in the range of 2,000 deg/s^2, with peaks of 6,000 deg/s^2 (List, 1983). Using helmet-mounted accelerometers, List developed both linear (constant velocity) and nonlinear methods for prediction. Researchers in the lab continued to study the effective limits of motion prediction for their current hardware testbed (Smith, 1984). While none of this work employed a Kalman filter, the growing importance of flight simulators to the U.S. Air Force meant that image generation, head tracking, and motion prediction were becoming hot topics. Among the various individuals studying the issues was Captain Robert Rebo, a graduate student at the U.S. Air Force Institute of Technology, Wright-Patterson Air Force Base. As part of his graduate work, Rebo began exploring the use of the Kalman filter for head-motion prediction.

Before moving on, it is important to note that during this time VR as a general paradigm was being actively explored by several others outside the U.S. Air Force, including the University of Utah, the NASA Ames Research Center, and the University of North Carolina at Chapel Hill (UNC). At UNC, under the direction of Henry Fuchs, graduate student Gary Bishop was working on methods for "self tracking" of a VR user—estimating full 6DOF head pose with a user-worn optical sensor cluster which looked outward at beacons (active sources or natural features) in the user's environment (Bishop, 1984). While Bishop did not actually use a Kalman filter, he was aware of earlier motion-tracking work such as H.J. Woltring's work (Woltring, 1974),

where the Kalman filter was used. In his Ph.D. dissertation Bishop stated, "A filter that uses information about the past position of the cluster and restrictions on its possible motions (e.g., a Kalman filter) could be used to allow proper operation for short periods with fewer than seven beacons visible but for reliability and accuracy the system must be designed so that seven or more beacons are visible essentially all the time" (Bishop, 1984).

In his dissertation Bishop also articulated a circular relationship between human motion and estimation algorithm complexity, namely, the simpler the estimation algorithm, the smaller the user motion between measurements, allowing for a simpler algorithm. Bishop recognized that the elegant, simple recursive nature of the Kalman filter made it very attractive in that sense.

4.2 Enter the Kalman Filter (1988–1994)

The first published account of the use of a Kalman filter in the context of VR appears to be Rebo's master's thesis (Rebo, 1988). There exists the possibility that it was being used on the VCASS system (see above) simultaneously or even prior to Rebo's work; however, it is only mentioned as part of work in development in Haas (1983). In any case, the available literature offers no clear link to the work. Rebo's work at the time was primarily in the area of HMDs. He built a prototype for the U.S. Air Force Super Cockpit project, employing Sharp miniature LCD displays, custom optics, and a Polhemus magnetic tracking system. While considering options for head-motion prediction, Rebo examined previous prediction work such as that by List (1983); however, recognizing that the Kalman filter was in use in other military systems (for example, to keep lasers on targets), he suggested that it could similarly be used for an HMD. He implemented such predictive HMD tracking using a Kalman filter on the full 6 DOF estimates from the Polhemus system. Nine months later Rebo and Phillip Amburn presented an update on the work at an SPIE conference on HMDs (Rebo & Amburn, 1989).

While it is a subtle point, it is interesting that Rebo and others in subsequent years seemed to consider the Kalman filter primarily as a tool for improved *motion*

prediction (by way of the time update step), as opposed to a general means for improved filtering of the 6 DOF pose estimates, which would then lead to improved prediction using any means—even outside the Kalman filter. On the other hand, given that researchers did not have access to the low-level measurements of the Polhemus system, perhaps there was not much else they could do. In addition, prior researchers had already articulated doubts about predicting more than a few image generation steps ahead in time, and thus the relatively short-term prediction inherent in the Kalman filter probably made the most sense.

A good illustration of this phenomenon is the characterization of Rebo's work by Robert Albrecht in the summer of 1989. In his M.S. thesis, Albrecht stated the following.

> The Kalman filter predictor did provide enhanced image stabilization for slower head movements, but did not significantly correct for lags at faster head movements. The Kalman filter predictor also allowed the image to "swim" at the end of quick head movements. Rebo recommended using a higher order Kalman filter implemented on a dedicated processor being continuously polled for the next look direction. (Albrecht, 1989, p. 34)

As an alternative, Albrecht experimented with several linear finite impulse response filters, in particular a variation of the least-mean-square algorithm presented by Woodrow and Stearns in 1985. As happens even today, it seems there was a tendency to attribute poor performance to the Kalman filter, rather than to the structure and parameterization of the process and measurement models, or the appropriateness of the use of the Kalman filter at all, given the nature of the measurements.

In any case, Rebo's work was discovered by researchers in the more general VR community, apparently spawning a new branch of work in general head tracking beyond flight simulators. The first example of such work appears to be work on temporal-spatial realism in VR published by Jiandong Liang, Chris Shaw, and Mark Green in 1990–1991 (Liang, Shaw, & Green, 1991). This work was solely aimed at improving the pose estimates from the Polhemus Isotrak magnetic tracker. The

authors characterized the raw Isotrak estimates as exhibiting noise and delay, and described the effect on the user as a temporal-spatial distortion. Through experiments they concluded that users were most perceptive to jitter in the position estimates and latency in the orientation estimates. Noting that jitter perpendicular to the line of sight was more noticeable than in other directions, and that most human motion is along the line of sight (and therefore filter-induced latency in that direction would be more noticeable), they developed an anisotropic low-pass filter for the position estimates.

To address the latency in the orientation estimates, they used a Kalman filter to facilitate motion prediction. The authors used four independent Kalman filters for the four elements of the orientation quaternion, and re-normalized the quaternion outside the filter. The authors employed a Gauss-Markov process model using first and second derivatives, and used the time constant β to model the limits of head velocities and accelerations. The authors noted that this model was more appropriate than the random-walk models used earlier by Rebo and Amburn, which they claimed suffered from excessive overshoot. Because the Polhemus Isotrak ran at 20 Hz, they used a 50 ms step size for the Kalman filter. The authors measured 151 ms of delay in the Isotrak estimates (not including rendering), and so experimented with tuning their filter for prediction three to 10 steps ahead. In the end they decided on three steps, in other words 150 ms.

Apparently around the same time, Martin Friedmann, Thad Starner, and Alex Pentland in the MIT Vision and Modeling Group were pursuing the use of the Kalman filter for tracking and predicting the motion of drumsticks for a project to interactively control computer-generated drums, bells, or strings, accompanied by synthesized sounds. To track the drumsticks, they used a Polhemus Isotrak system, and again latency was the dominant issue in terms of synchronizing the live physical (real) drumming with the graphics and sounds of the virtual drums. This work resulted in two publications (Friedmann, Starner, & Pentland, 1992a, b) which followed Liang, Shaw, and Green's work (Liang et al., 1991) by two to four months. While their later publication references Liang, Shaw, and Green's work when

talking about the orientation filtering, their earlier publication makes no mention of previous work by Liang, Shaw, and Green, or Rebo. From the articles it appears that their use of the Kalman filter grew out of other computer vision work at MIT at the time, and they only became aware of the other VR work later.

In any case, synchronizing the motion of real drumsticks with virtual drums must have been exceedingly challenging since it involves sight, sound, haptic feedback, and proprioception. Friedmann, Starner, and Pentland used a Kalman filter to estimate the positions and velocities of the tips of the drumsticks using measurements from the Isotrak sensors. Interestingly, they did not use a Gauss-Markov process model as Liang, Shaw, and Green did; however, instead they used a random walk with experimentally tuned noise parameters. Even with their best tuning efforts, they observed overshoot of the predictions, which caused users to alter (slow) their motion unnaturally. Friedmann, Starner, and Pentland noted that the overshoots were not a problem as long the drumstick was far from the drum head, since it only exaggerated the user's motion; however, they were a problem when near the drum head, since the virtual drumstick would pass through the virtual drum head. Their solution was to implement a multiple-motion hypothesis scheme with multiple Kalman filters, and to choose the solution with the greatest likelihood, given the recent measurements and some rules for likely motion. Aside from their simple process model, this work appears to be relatively sophisticated compared to other VR work at the time. Unfortunately the work was cut short when Friedmann died unexpectedly in 1995.

Also at MIT, Ali Azarbayejani was working with Thad Starner, Brad Horowitz, and Alex Pentland on what they called *visually controlled graphics* (Azarbayejani, Starner, Horowitz, & Pentland, 1993). Their goal was to develop a head-tracking system that used cameras and passive computer vision, thus freeing the user from the encumbrances of the Polhemus Isotrak sensor wires. While the working volume was limited compared to that of an Isotrak, this work was noteworthy in several respects. First, it appears to be the first nexus of VR, computer vision, and the Kalman filter. In addition, it ap-

pears to be the first work in VR that used a Kalman filter on low-level sensor measurements (features in camera images) as opposed to complete pose estimates from another system such as the Polhemus Isotrak. Finally, it appears to be the first work that employed an *extended* Kalman filter, necessitated by the nonlinear projection involved in cameras.

The measurements Azarbayejani, Starner, Horowitz, and Pentland used were points in the 2D camera images where the image intensity surface had a large Hessian for some threshold, which corresponds to peaks, saddle points, and pits in the intensity surface. The authors modeled the user's head with a simple ellipsoid with presumed visible 3D feature points on the surface. The authors initialized the 3D surface points using projections of the initially seen (in the camera images) 2D feature locations, and then recursively estimated the overall pose of the ellipsoid using the 2D observations of the 3D surface points. In a separate publication they also demonstrated the simultaneous estimation of the distances of the 3D points from the center of the ellipsoid, thus simultaneously estimating the pose and the structure of the head (Azarbayejani, Horowitz, & Pentland, 1993).

Around this same time, apparently inspired by the recent work of Friedmann, Starner, and Pentland at MIT, and Liang, Shaw, and Green, Satoru Emura and Susumu Tachi from the Research Center for Advanced Science and Technology at the University of Tokyo had the idea to use the Kalman filter to fuse low-level gyro measurements with Polhemus tracker estimates to achieve a higher estimation rate and lower latency than the Polhemus tracker alone. Emura and Tachi formulated the process and measurement models to do this, and developed a novel method for estimating the system delay by performing correlation with a reference signal from a mechanical link-type system. Emura and Tachi formally presented their idea and experimental results in July of 1994 (Emura & Tachi, 1994).

The final body of work to close out this period of introduction of the Kalman filter to VR is that of Ron Azuma and Gary Bishop at UNC at around the same time. On the heels of the earlier-mentioned self-tracking work by Bishop in the early 1980s (Bishop, 1984), re-
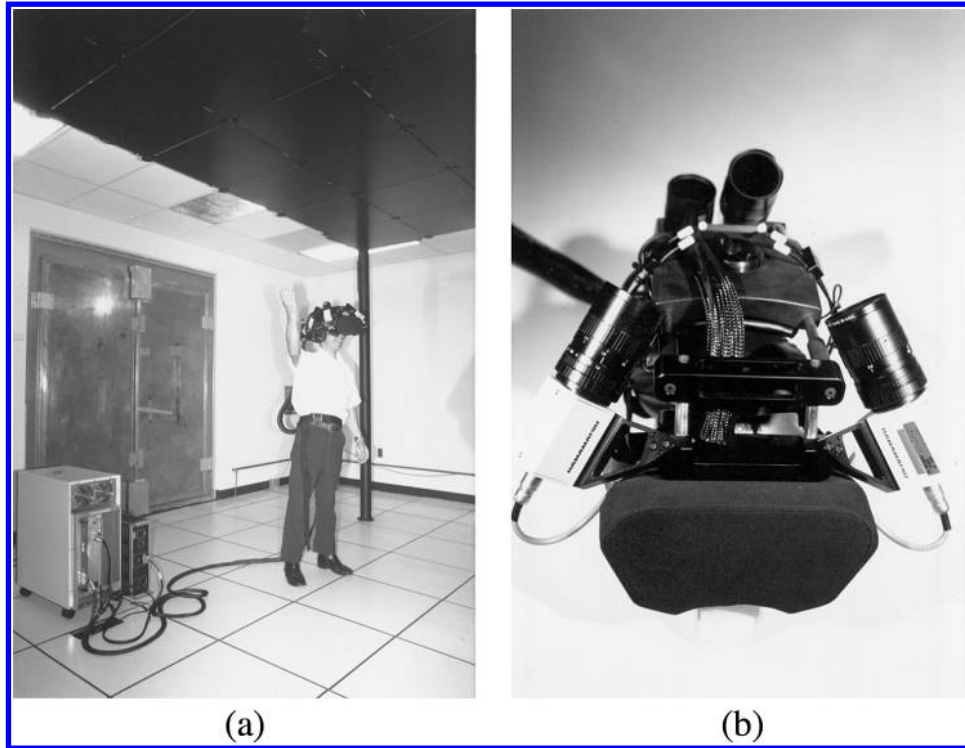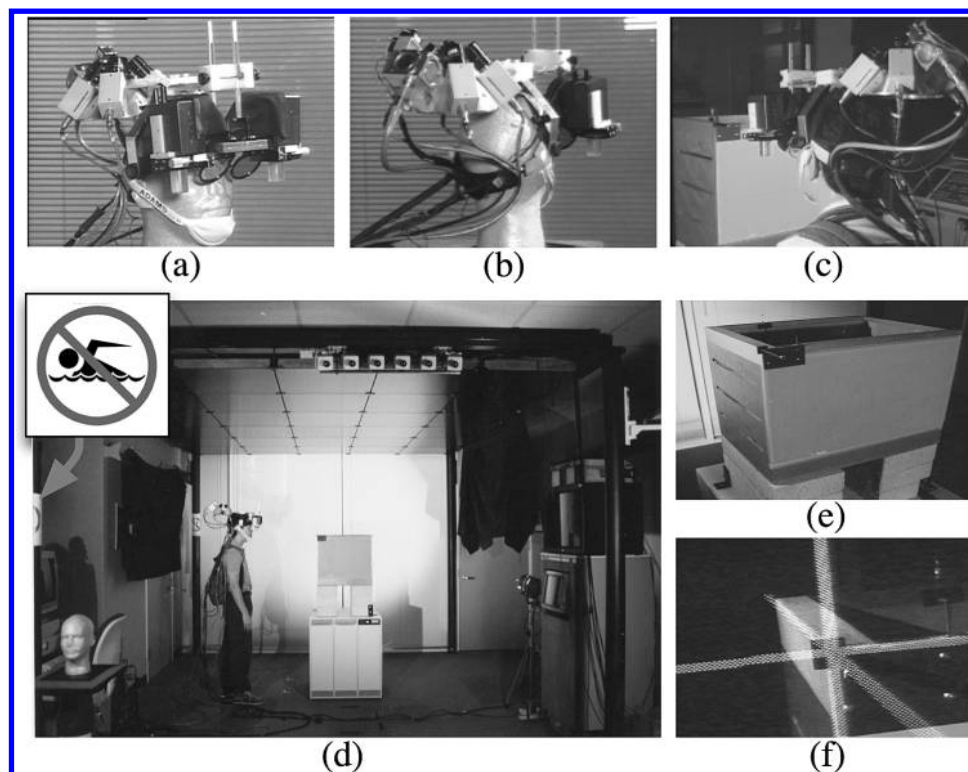
Figure 3. *The University of North Carolina's active optoelectronic head-tracking system (c. 1992). The system was composed of a specially machined ceiling with a 2D array of infrared LEDs embedded in it as shown in (a), narrow-band infrared passing lateral-effect photodiode optical sensors mounted on the head-mounted display as shown in (b), and a backpack containing some signal processing electronics. Because the photodiodes measure the centroid of all light impinging on the sensor, only one LED could be illuminated at a time. As such, the system operated by sequentially illuminating LEDs in the ceiling, collecting the corresponding 2D measurements from the head-mounted photodiodes, and then using photogrammetric techniques to estimate the head pose. The approach was often likened to navigating by the stars. Images courtesy of UNC-Chapel Hill.*

searchers at UNC began working on an active optoelectronic head-tracking system (Ward, Azuma, Bennett, Gottschalk, & Fuchs, 1992). The system was comprised of a specially machined ceiling with a 2D array of infrared (IR) light-emitting diodes (LEDs) embedded in it as shown in Figure 3a, and narrow-band IR passing lateral-effect photodiode (LEPD) optical sensors mounted on the HMD as shown in Figure 3b. Because the LEPDs measure the centroid of all light impinging on the sensor, only one LED could be illuminated at a time. As such, the system operated by sequentially illuminating LEDs in the ceiling, collecting the corresponding 2D measurements from the head-mounted cameras, and then using photogrammetric techniques to

estimate the HMD (head) pose from the batch of measurements. The approach was often likened to navigating by the stars. The system, which was demonstrated in the Tomorrow's Realities gallery at the 1991 conference of the Association of Computing Machinery (ACM) Special Interest Group on Graphics and Interactive Techniques (SIGGRAPH), had a scalable working area that measured 10 ft by 12 ft, a measurement update rate 20–100 Hz with 20–60 ms of delay, and a resolution specification of 2 mm and 0.2° (Wang et al., 1990). This system covered the widest area and offered the highest performance of its day.

From the very beginning, the UNC researchers repeated the mantra of "no swimming"—an allusion to

Figure 4. *Images related to 1994 work by Azuma and Bishop at the University of North Carolina on using a Kalman filter to fuse inertial and active-optical measurements to improve the dynamic visual registration in a see-through head-mounted display (HMD). The custom HMD and associated inertial and optical sensors are shown in (a)–(c). The custom-built inertial pack can be seen in the upper left of (b) and the upper right of (c). Azuma is shown using the tracking system in (d). An inset in (d) shows a "no swimming" sign used to remind everyone of the desire to reduce the visual effects of excessive latency. The test setup is shown in (c) and (e), and a view through the optics of the see-through HMD in (f), where a set of three virtual white lines are properly aligned with the corner of the test rig. Images reproduced from Azuma (1995), with permission from UNC-Chapel Hill.*

their goal of eliminating the then-common problem of visual effects resulting from a combination of delay, overshoot, and the user's attempts at correction. As mentioned earlier, the challenges are even greater for *augmented* reality (AR), where the computer graphics images are blended with a user's view of the real world. This challenge became a very concrete problem for UNC's Henry Fuchs and Andrei State, who were attempting to use the tracker for AR-assisted amniocentesis, a joint project with physicians at UNC Memorial Hospital. The idea was that during an amniocentesis procedure, the physician would use an ultrasound probe and an HMD, and thus be able to see virtual 3D imag-

ery of the baby superimposed on the mother while he or she inserted the (not small) needle. Basically they wanted to give the doctor "X-ray vision." For this purpose they developed a special *see-through* HMD using half-silvered mirrors to optically blend the real and computer-generated imagery for the user. The custom HMD and associated tracking sensors are shown in Figure 4a–c. As it turns out, even the respectable 20–60 ms of tracking delay, when combined with rendering and other system delays, resulted in nontrivial misregistration between the needle and the 3D imagery, with only seemingly small head motions.

As in the past, the researchers attacked the problem

by attempting to reduce all of the delays to their minimum, and employing head-motion prediction. The issue here was that, similar to the Polhemus system, the optoelectronic tracking system offered no direct measurement of the head velocities and accelerations. Instead, the derivatives had to be estimated from the position and orientation estimates. The derivative estimates were inherently delayed and noisy.

To address the prediction problems, Ron Azuma, Gary Bishop, and Vernon Chi began to explore the addition of inertial sensors to the existing optoelectronic system. The issue then became one of how to fuse the optical and inertial measurements. Bishop had been aware of the Kalman filter for years—at least since Bishop (1984)—and was aware of the work by Rebo, Liang, and Friedmann. He suggested to Azuma that it was likely to be the best method for fusion. Azuma and Bishop subsequently modified the see-through HMD, adding a custom-built inertial pack consisting of three Systron Donner QRS-11 angular rate gyroscopes and three Lucas NovaSensor NAS-CO26 linear accelerometers. The inertial pack can be seen in Figure 4 in the upper left of (b) and the upper right of (c).

Azuma and Bishop implemented a Kalman filter to fuse the position and orientation estimates from the ceiling tracker with the derivatives from the inertial sensors, and as had been done in the past using the Kalman filter process model to predict (extrapolate) the head pose into the future. Azuma and Bishop developed new interactive procedures for estimating the viewing transforms for the HMD—which, as Holloway argued, are critical to reducing static registration error (Holloway, 1995)—and new methods for automatically estimating the pose of the inertial pack in the coordinate frame of the HMD. Azuma is shown using the tracking system in Figure 4(d). An inset in (d) shows a "no swimming" sign used to remind everyone of the goal. The AR test setup is shown in (c) and (e), and a view through the optics of the see-through HMD in (f), where a set of three virtual white lines are properly aligned with the corner of the test rig. Azuma and Bishop presented this work at ACM SIGGRAPH 1994, where they reported that "[o]n average, prediction with inertial sensors produces errors 2–3 times lower than prediction without

inertial sensors and 5–10 times lower than using no prediction at all" (Azuma & Bishop, 1994, p. 197). By that time, ACM SIGGRAPH had become the premier venue for presenting VR-related work. Azuma and Bishop's work has been widely referenced since that time, in virtually all VR-related papers using the Kalman filter. The Kalman filter had truly arrived in VR.

4.3 Widespread Use of the Kalman Filter

In the summer of 1995, Azuma and Bishop followed their SIGGRAPH 1994 paper with a SIGGRAPH 1995 paper on a frequency-domain analysis of head-motion prediction (Azuma & Bishop, 1995). In addition to describing an analytical method for comparing head-motion prediction approaches, including those using the Kalman filter, they arrived at the conclusion that accelerometers might be the most valuable inertial sensors to use for human head-motion tracking. This publication also appears to be the first place where error analysis was done in the screen space of the HMD. Screen space (eye space) is where all error ultimately manifests itself in any VR system, and hence is arguably the most appropriate space to do error analysis. After finishing his Ph.D. at UNC in 1995, Azuma joined the Hughes Research Lab (HRL) and then the Nokia Research Center Hollywood. He continues to work on Kalman filter-based tracking for VR and AR, publishing widely in the field.

Back in 1992, So and Griffin had introduced the idea of *image deflection* for HMDs (So & Griffin, 1992). The basic idea was to render the HMD imagery using the best/latest head pose prediction from the tracking system, then query the tracking system again just prior to displaying the final image after the rendering completes, and shift the image horizontally and vertically an appropriate amount in the display frame buffer to reduce the angular projection of the effects of head-pose error. This last-minute update to the imagery is possible because shifting the rendered image can be done relatively quickly compared to the rendering itself. In 1995 Tomasz Mazuryk and Michael Gervautz published a method for "Two-step prediction and image deflection"

which used a Kalman filter framework to estimate the current head pose, and a separate (external to the Kalman filter) prediction process (Mazuryk & Gervautz, 1995). Even though the head position and velocity are the integrals of the head velocity and acceleration, respectively, the authors used two completely independent Kalman filters to estimate the head position and the derivatives. Their rationale was that having independent filters allowed them to tune their process models independently. The authors stated they were able to achieve better estimation results with this two-step process than a single filter. The authors then employed subsequent prediction and deflection steps, not unlike So and Griffin.

Around the time Azuma was leaving UNC for HRL, Welch (the author) began working with Bishop on tracking research. Welch and Bishop had really begun to focus on the Kalman filter as the best approach to on-line, real-time head tracking for VR. Coming from NASA's Jet Propulsion Laboratory, Welch had noted that spacecraft typically navigated using a combination of inertial sensors and optical devices (star and sun trackers), and he wondered whether the same complementary approach would be appropriate for head tracking in VR. He had been pursuing the idea as a Ph.D. topic, then later switched topics, given the promise of an earlier completion.

Welch articulated the hybrid inertial/optical ideas, including the use of a complementary error-state Kalman filter, in a 1995 technical report (Welch, 1995). This publication led to a DARPA proposal by Bishop, Welch, and Chi at UNC, and eventually a project award to a team which included UNC, HRL (Azuma), the University of Southern California (Neumann), and Raytheon. The project, "Geospatially-registered information for dismounted soldiers," was aimed at developing tracking, rendering, and user interface technologies for outdoor AR for dismounted warfighters. The idea was that the warfighter would be able to to see virtual indications of friends and foes, navigational aids, and objects otherwise occluded by buildings or the landscape. The approach was to use a Kalman filter to fuse measurements from multiple sensors, including—most importantly—measurements of features in the imagery from

the cameras that were providing the warfighters with their view of the real world. This hybrid of multiple motion sensors and closed-loop vision-based tracking set the preliminary groundwork for most of the subsequent outdoor AR work, including years of subsequent work by several of the team members.

On April 12, 1997, researchers at UNC obtained the very first results from the next-generation version of the optoelectronic tracking system described earlier and shown in Figure 3. The new system, known as the Hi-Ball tracking system, operated similarly to the original system in that it used ceiling-mounted LEDs and user-worn LEPDs; however, it included dramatic improvements in both hardware and software. For example, the relatively bulky sensor rig shown in Figure 3b, along with some associated electronics contained in a user-worn backpack, were reduced to the form factor shown in Figure 5a. By embedding most of the LEPD signal processing circuitry in the HiBall, the developers were able to achieve measurement rates of up to a theoretical maximum of 5,000 single LED measurements per second.

At the same time, the head-pose estimation method was converted from the relatively slow batch photogrammetric approach used in the original system to an extended Kalman filter using sequential LED measurement processing. While the LED-LEPD measurement system inherently restricted any algorithm to sequential measurements, and there had been previous work in bearings-only estimation, the notion that one would intentionally refine the 6 DOF pose estimates with sequential 2D measurements, rather than sequentially assembling a batch of measurements that sufficiently constrained a closed-form solution, was new to the VR community. Welch and Bishop presented the work at ACM SIGGRAPH 1997, calling it single-constraint-at-a-time (SCAAT) tracking (Welch & Bishop, 1997). The SCAAT approach was inherently aimed at taking advantage of the aforementioned circular relationship between human motion and estimation algorithm complexity. Compared to all previous batch methods, one Kalman filter cycle with a single 2D measurement had dramatically fewer floating point instructions, which meant that the a priori state estimates could be updated
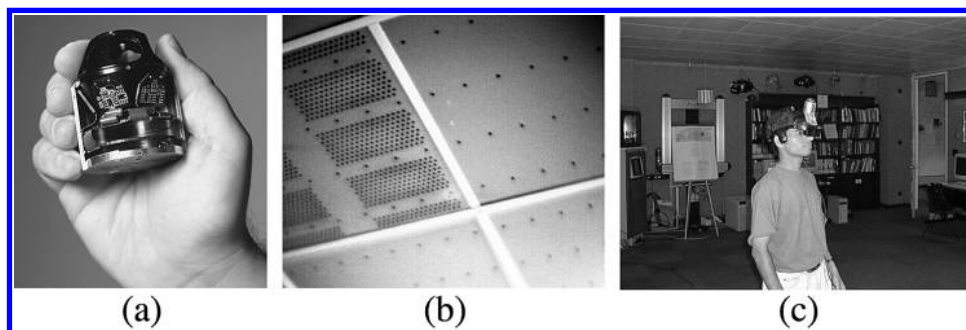
Figure 5. *The first prototype of the HiBall optoelectronic tracking system developed at the University of North Carolina (c. 1997). This system is the second generation version of the optoelectronic tracking system shown in Figure 3. The HiBall system operated similarly to the original system in that it used a user-worn optical sensor unit shown in (a) to observe ceiling-mounted LEDs shown in (b); however, it included dramatic improvements in both hardware and software. For example, the relatively bulky sensor rig shown in Figure 3(b), along with some associated electronics contained in a user-worn backpack, were reduced to the form factor shown in (a). The precision LED ceiling panels shown in Figure 3(a) were reduced to the simple drop-in versions shown in (b). In (c), graduate research assistant Pawan Kumar is shown using the prototype system. Images reproduced from Welch et al. (2001), with permission from MIT Press.*

very quickly. Hence the predictions associated with the filter's time update would usually be quite good; thus the linear approximations of the extended Kalman filter worked well.

By taking advantage of the high measurement rate, and carefully modeling the expected dynamics of the user, Welch and Bishop were able to augment the filter state (head pose and derivatives) with the location of the LED being used for each measurement. By effectively using a separate Kalman filter for each LED, they were able to estimate the user's 6 DOF head pose while simultaneously and automatically estimating a map of all LED locations. This approach enabled the use of much simpler drop-in ceiling panels with embedded LEDs, as shown in Figure 5b. The resulting system, which is shown in use in Figure 5c, produced estimates at up to 3 kHz, with around 1 ms of latency, on the order of a few tenths of a millimeter of position resolution and accuracy, and on the order of hundredths of a degree of rotation noise and accuracy. A commercial version of the HiBall system is now sold by 3rdTech, Inc. as the 6 DOF HiBall-3100.

While learning about the Kalman filter during this period, Welch and Bishop grew to appreciate both the

relatively general introduction given by Peter Maybeck in Chapter 1 of Maybeck (1979), and the more detailed presentations given in the subsequent chapters of that and several other books. Welch and Bishop felt that writing something that fell between the existing explanations would both help them learn and potentially be useful to others. As such, in 1995 they wrote their own "Introduction to the Kalman filter" technical report (Welch & Bishop, 1995), which has been widely referenced, and even translated into Chinese. Similarly, they recognized that there was no single home for the Kalman filter on the web—no place where one could find information about books, papers, and software, much less information on R. E. Kalman himself. This recognition led them to create a Kalman filter web site (Welch & Bishop, 2008). The site has had almost 800,000 visits over the past 11 years (roughly 200 hits per day) from all over the world.

Back in 1993, a then relatively unknown Eric Foxlin (né Eric Fuchs) had finished his M.S. degree at MIT under the direction of Nathaniel Durlach. Foxlin, who had received a Link Foundation fellowship during his final year, remained at MIT until 1996, leading the Head-Tracker Development Project in the Sensory

Communications Group of the Research Laboratory of Electronics. In 1996 he left MIT and formed InterSense, Inc., which today is one of the dominant VR tracking-system companies. More important to the context of this paper, virtually every product offered by InterSense, and virtually every paper published by Foxlin since 1996, has involved some use of a Kalman filter.

Three examples of Foxlin's work are important to mention in the context of this article. The first example is his 1996 work on a 3 DOF (orientation only) head-tracking system that used a complementary separate-bias Kalman filter to fuse measurements from angular rate gyros, inclinometers, and a fluxgate compass (Foxlin, 1996). Referring to the 1965 work by B. E. Bona and R. J. Smay on resetting a ship's inertial navigation system (Bona & Smay, 1965), Foxlin presented a means to use a complementary error-state Kalman filter to independently estimate the gyro drift, which he would reset when the user was deemed still. The work was significant in that it took advantage of the rapidly improving micro-electro-mechanical systems (MEMS) based inertial sensor development to create a small, rigid, and low-latency orientation tracking system. An early prototype from MIT is shown in Figure 6a. The work led directly to the first commercial product by InterSense, the 3DOF InterSense InertiaCube, which is still available today.

The second example of Foxlin's work is his work with Michael Harrington and George Pfeifer on a hybrid inertial-acoustic (time-of-flight) 6 DOF tracking system they called the Constellation (Foxlin, Harrington, & Pfeifer, 1998). In some ways the approach they used was similar to the UNC optoelectronic systems (the Hi-Ball and earlier) in that the Constellation employed small acoustic transmitters throughout the working volume, and microphones on the user's head. And while acoustic-range measurements are relatively inaccurate compared to the angular measurements of LED sightings, the user-worn acoustic components are relatively small, and the fusion with measurements from a user-worn inertial pack containing both angular rate gyros and linear accelerometers resulted in good performance over a wide area. Foxlin again employed a complementary error-state Kalman filter that estimated the error in
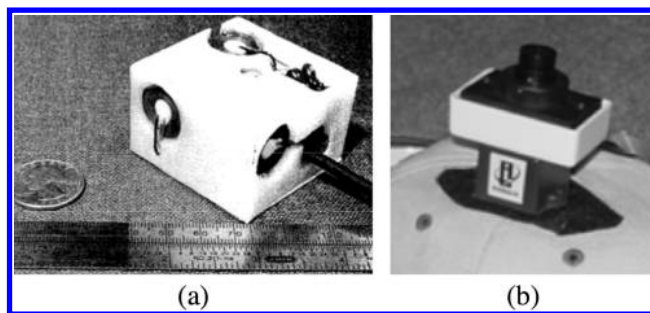
Figure 6. *Prototype sensor units by Foxlin at MIT and InterSense. (a) A 1996 prototype of a 3 DOF (orientation only) sensor unit that included microelectronic machine type angular rate gyros, inclinometers, and a fluxgate compass. The associated head-tracking system used a complementary separate-bias Kalman filter to fuse the individual device measurements from the sensor unit (Foxlin, 1996). (b) A 2003 prototype of a sensor unit that incorporated an InterSense orientation sensor, combined with a camera and specialized embedded processing to find and track 2D image features corresponding to printed patterns placed in the environment. The associated head-tracking system used a generalized Kalman filter-based architecture for simultaneous localization, auto-calibration, and map-building (Foxlin, 2002a). The work led directly to the most recent commercial product by InterSense, the 6 DOF InterSense IS-1200. Images (a) and (b) reproduced from Foxlin (1996) and Foxlin and Maimark (2003) respectively, with permission from IEEE, © 1996 and © 2003.*

the integrated 6 DOF inertial signals, using the period acoustic range measurements in a SCAAT fashion as in Welch and Bishop (1997). The work led directly to the second major commercial product by InterSense, the 6 DOF InterSense IS-600.

The final example of Foxlin's work is the hybrid inertial-vision (computer vision) development he undertook with Leonid Naimark (Foxlin & Naimark, 2003), enabled in part by some recent work he had done on a generalized Kalman filter-based architecture for simultaneous localization, auto-calibration, and map-building (Foxlin, 2002a). Foxlin's goal was to eliminate the need for any active beacon infrastructure (acoustic or optical) by instead using black and white encoded patterns that could be printed from virtually any personal computer and then attached to the ceiling with cellophane tape. The fundamental idea for an inertial-vision hybrid using

a complementary error-state Kalman filter was articulated by Welch (1995) and pursued by Matt Cutts, Welch, and Bishop at UNC in the late 1990s. The researchers were attempting to control inertial drift by using natural features seen from a camera pointed toward the ceiling. During that time, Foxlin visited UNC, and later told Welch that the UNC work inspired him to pursue something similar. Two significant aspects of Foxlin and Naimark's 2003 work were the use of printed encoded patterns such as in Artoolkit (2005) on the ceiling, and the use of specialized embedded (in the camera) processing to find and track 2D image features at a relatively high rate of speed. An early prototype is shown in Figure 6b. The work led directly to the most recent commercial product by InterSense, the 6 DOF InterSense IS-1200. Around the same time, Foxlin and Welch co-authored a paper in IEEE Computer Graphics Applications on motion tracking for virtual and augmented reality (Welch & Foxlin, 2002).

In 2002, Nick Vallidis completed his Ph.D. at UNC under the direction of Gary Bishop. Vallidis was the third Ph.D. student of Bishop's to employ the Kalman filter for tracking in VR. Vallidis's work involved using the Kalman filter for a novel form of acoustic tracking. Acoustic signals had been in use for tracking in VR since the very beginning, when Ivan Sutherland used them to track a user's head (Sutherland, 1968). While both continuous wave and time-of-flight methods had been previously explored, Vallidis explored the use of spread-spectrum acoustic signals. His approach was to transmit broadband pseudorandom codes from small transducers (speakers) on the moving target, and to correlate the known codes with the signals received by way of microphones on a stationary reference. Unlike time-of-flight or continuous wave methods, the broadband nature of the codes resulted in robustness to partial occlusions, since lower frequencies could pass around partially occluding objects. In particular, this made the approach attractive for near-body limb tracking—tracking hands, elbows, knees, and feet with respect to a person's body. Vallidis used a Kalman filter for range measurements, tracking the point of maximum signal correlation, while estimating Doppler and other effects. Using multiple Kalman filters, he estimated range (1 DOF), position

(3 DOF), or complete pose (6 DOF). Because the pseudorandom signals covered most of the audible spectrum with equal energy, they had a white-noise sound that led to the method being dubbed Whisper tracking (Vallidis, 2002).

Today the most common use of the Kalman filter for VR or AR tracking seems to be related to the fusion of computer vision measurements with other sensors, primarily inertial, as originally described in Welch (1995). The emphasis on computer vision for tracking follows a larger trend toward the merging of computer graphics and computer vision. This trend is perhaps not surprising, since much of computer graphics is aimed at rendering realistic scenes, which these days includes image-based rendering, while much of computer vision is aimed at interpreting real scenes, which sometimes includes predicting them using graphics models. At conferences such as IEEE Virtual Reality (IEEE VR) and the International Symposium on Mixed and Augmented Reality (ISMAR) over the past years, there are quite a few examples of people using cameras, inertial devices, global positioning systems, and Kalman filters for VR and AR. For example, Stephen DiVerdi and Tobias Höllerer at the University of California at Santa Barbara have been working on paradigms whereby the user specifies nearby landmarks that are treated as measurements (DiVerdi & Höllerer, 2007). Similarly, they are working on incorporating constraints from 2D road maps of the area.

Another recent example is work being done by Gabriele Bleser and Didier Stricker at the Fraunhofer Institute for Computer Graphics in Darmstadt, Germany. At IEEE VR 2008 they presented their latest work on a Kalman filter based inertial-vision device they call GroundCam. The GroundCam, shown in Figure 7a, appears to be similar to the inertial-vision units built by Foxlin in Figure 6c. However, Bleser and Stricker's work is aimed at tracking without special visual markers, instead using natural features in the environment—corners of objects, edges, and other visible features. The approach relies on a 3D model of the scene to predict the locations and appearances of the features by rendering the model using the prediction data of the Kalman filter. By also modeling and rendering the lighting in
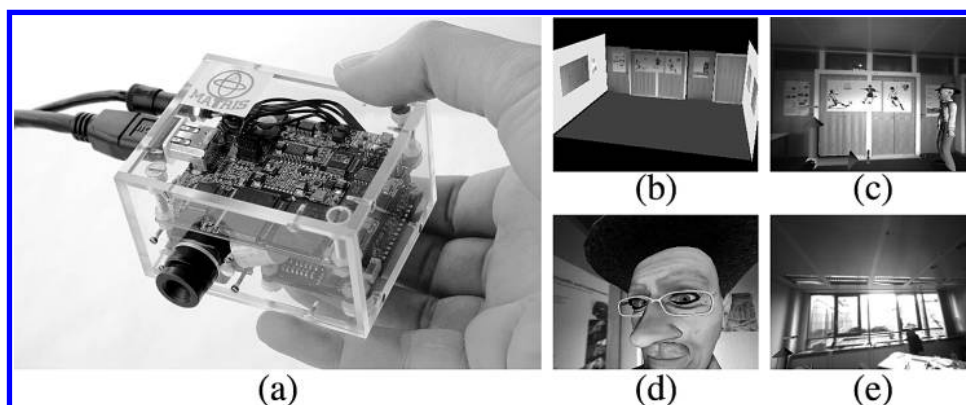
Figure 7. *Images related to work by Gabriele Bleser and Didier Stricker at the Fraunhofer Institute for Computer Graphics in Darmstadt, Germany, on a Kalman filter based inertial-vision tracking system they call GroundCam. The GroundCam system was designed to be used for augmented reality, where three-dimensional computer graphics are superimposed on camera imagery of the real world. Subfigure (a) shows the current prototype of the device, which includes inertial sensors and a camera. While this appears to be similar to the inertial-vision units built by Foxlin and InterSense in Figure 6c, the GroundCam is aimed at tracking without special visual markers, instead using natural features in the environment such as corners of objects, edges, and other visible features. Their approach relies on a 3D model of the scene to predict the locations and appearances of the features by rendering the model using the prediction data of the Kalman filter (Bleser & Stricker, 2008). Images reproduced from Bleser and Stricker (2008), with permission from the authors.*

the environment, they improve the robustness and accuracy of the feature image localization (Bleser & Stricker, 2008).

The right of Figure 7 shows some example results. Subfigure (b) shows a 3D computer graphics model of a room, with lighting. The graphics model is used, with the pose estimates from the Kalman, to predict feature locations and appearances, which are the camera measurement predictions for the Kalman filter. Subfigures (c)–(e) show some images from the GroundCam camera shown in subfigure (a), with corresponding and properly aligned computer graphics overlaid. The idea is that the virtual objects and characters should seem as if they are in the room.

Beyond head, hand, and device tracking, there has been some interesting recent Kalman filter based motion capture work, some of which has found its way into commercially available products. Motion-capture systems are used to obtain dynamic kinematic models of moving humans (for example, for motion studies or movie making). A historically popular approach is opti-

cal motion capture. The subjects wear black Lycra suits with golfball-sized retroreflective spheres attached all over. Special infrared (IR) cameras with IR ring lights illuminate the scene and measure the 2D image coordinates of the spheres, and in software the systems estimate the dynamic posture of the moving bodies. However, optical motion capture is now having to make room for inertial systems (all using Kalman filters) which are eliminating the need for infrastructure, allowing the participant's motion to be captured beyond the lab, outdoors—virtually anywhere.

Interesting work on this was originally pursued by Eric Bachmann at the U.S. Naval Postgraduate School in 2000. Bachmann built custom small devices which included magnetometers and inertial sensors (linear accelerometers and angular rate gyros). He attached the devices to various points on the user's body, and then using Kalman filters and a kinematic model of the body (for prior constraints), he was able to record human motion without any cameras (Bachmann, 2000). Around that same time, Shin, Lee, Shin, and Gleicher

(2001) and Tak, Song, and Ko (2002) were using the Kalman filter to aid in on- or off-line interpretation of motion capture data. Luinge and Veltink (2005) used the Kalman filter to help estimate the orientation of human body segments using inertial sensors but no magnetometers. To obtain an absolute reference for pitch and roll (not heading) they measured the gravity vector using accelerometers, and then combined this with the integrated outputs of gyroscopes using an *error-state* (or *complementary*) Kalman filter. In this way the accelerometers served as an aid to the system, by providing an independent measure of the gravity vector, which is used to help control orientation drift. This is similar to some earlier work by Foxlin (1996). More recently, Daniel Vlasic and his collaborators used a similar approach; however, they employed time-of-flight acoustic ranging to constrain the distances between the sensors, thus improving the robustness and accuracy of the captured motion (Vlasic et al., 2007).

Commercially, a company called Animazoo has developed an inertial motion-capture system that uses 3 DOF InertiaCube rotation sensors from InterSense mounted on each limb to estimate the dynamic articulated body models. Another company, called Xsens, has developed an inertial motion capture suit from the ground up, with custom inertial sensors. Xsens's Moven system uses Kalman filters to track the position and orientation of each body segment based on custom sensors that include angular rate gyros and linear accelerometers, enabling the capture of vertical and horizontal motion such as jumping and running.

5 Conclusions

The "Holy Grail" for researchers working on tracking for VR/AR still seems to be robust and accurate tracking outdoors, for augmented reality everywhere. Researchers around the world are working on 6 DOF position and orientation-aware computer interfaces that will support access to information embedded in or attached to the physical world all around us. From the laboratory, to the hallway, and beyond to parks and city sidewalks, individuals will some day see, hear, and interact with information that exists as an integral part of their immediate physical surroundings. The interfaces would range from office-based or portable display systems, to handheld and head-worn devices. A pose-aware projector would become an "information lamp" that projects relevant imagery directly onto the physical surroundings, while a pose-aware, handheld display would become a "magic lens" through which one sees information attached to the physical surroundings. Doctors will remotely assist others, workers and technicians will be guided in assembly and maintenance, blind people given gaze-directed aural sight, and deaf people visual hearing. Information and associated databases will be organized by physical location and time, allowing users to both store and retrieve past, present, and future information in the context of physical locality and direction of gaze. The Kalman filter will undoubtedly play a role in this vision, no matter what the underlying sources of signals.

In 1994, Frederick P. Brooks, Jr. became the first recipient of the ACM Allen Newell Award. His acceptance lecture, delivered at ACM SIGGRAPH 1994, was in part based on his famous 1977 "Computer scientist as toolsmith" article (Brooks, 1977). Brooks' characterization of computer scientists as toolsmiths was based on the notion that they "do not themselves directly satisfy human needs" but instead develop tools that "others use in making things that enrich human living" (Brooks, 1996, p. 62). Clearly the Kalman filter is such a tool. Brooks also stated that "[a] toolmaker succeeds as, and only as, the users of his tool succeed with his aid. However shining the blade, however jeweled the hilt, however perfect the heft, a sword is tested only by cutting. That swordsmith is successful whose clients die of old age" (Brooks, 1996, p. 62). For the VR community, the Kalman filter has proved to be a sword that remains sharp even after 50 years. Indeed, because it is employed in the most widely used VR tracking systems, it continues to impact all of the fields where VR is used. Examples are scientific visualization, simulation and training, medical and health-related applications, engineering, design, art, and entertainment. While certainly sensor and other technology improvements have been

invaluable, the Kalman filter arguably deserves significant credit for eliminating tracking as the dominant source of visual errors for such mainstream applications of VR.

Acknowledgments

The author would like to acknowledge the help obtained from various individuals, in particular in the way of personal recollections of events which in some cases took place 20 or more years earlier. The individuals include (alphabetically) Ron Azuma, Gary Bishop, Michael Deering, Steve Feiner, Mark Green, Tom Furness, Rory Stuart, and Brad Paley. The author would also like to thank the UNC-Chapel Hill Interlibrary Loan staff for their remarkable ability to obtain copies of obscure papers and to do so in a timely fashion.

References

Albrecht, R. E. (1989). *An adaptive digital filter to predict pilot head look direction for helmet-mounted displays.* Master's thesis, University of Dayton, Dayton, OH.

Artoolkit. (2005). Available at http://www.hitl.washington.edu/artoolkit/.

Azarbayejani, A., Horowitz, B., & Pentland, A. (1993). Recursive estimation of structure and motion using relative orientation constraints. *1993 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR '93,* 294–299.

Azarbayejani, A., Starner, T., Horowitz, B., & Pentland, A. (1993). Visually controlled graphics. *IEEE Transactions on Pattern Analysis Machine Intelligence, 15*(6), 602–605.

Azuma, R. T. (1995). *Predictive tracking for augmented reality.* Ph.D. thesis, University of North Carolina at Chapel Hill.

Azuma, R. T., & Bishop, G. (1994). Improving static and dynamic registration in an optical see-through HMD. *SIGGRAPH '94, Computer Graphics, Annual Conference on Computer Graphics & Interactive Techniques,* 197–204.

Azuma, R. T., & Bishop, G. (1995). A frequency-domain analysis of head-motion prediction. *SIGGRAPH '95, Computer Graphics, Annual Conference on Computer Graphics & Interactive Techniques,* 401–408.

Bachmann, E. R. (2000). *Inertial and magnetic tracking of limb segment orientation for inserting humans into synthetic environments.* Ph.D. thesis, The Naval Postgraduate School.

Bhatnagar, D. K. (1993). *Position trackers for head mounted display systems: A survey.* (Tech. Rep. TR93-010). University of North Carolina at Chapel Hill.

Bishop, G. (1984). *The self-tracker: A smart optical sensor on silicon.* Ph.D. thesis, University of North Carolina at Chapel Hill.

Bleser, G., & Stricker, D. (2008). Advanced tracking through efficient image processing and visual-inertial sensor fusion. *Proceedings of IEEE Virtual Reality Conference, 2008, VR '08.*

Bona, B. E., & Smay, R. J. (1965). Optimum reset of ship's inertial navigation system. *IEEE Transactions on Aerospace and Electronic Systems,* AES-2, 409–414.

Brooks, F. P., Jr. (1977). The computer scientist as toolsmith—Studies in interactive computer graphics. In B. Gilchrist (Ed.), *Information Processing 77, Proceedings of IFIP Congress 77,* 625–634.

Brooks, F. P., Jr. (1996). The computer scientist as toolsmith ii. *Communications of the ACM, 39*(3), 61–68.

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

Brown, R. G., & Hwang, P. Y. C. (1996). *Introduction to random signals and applied Kalman filtering: With MATLAB exercises and solutions* (3rd ed.). New York: Wiley & Sons.

DeRuyck, A. R., & Kuipers, J. (1973). *An electromagnetic position sensor.* (Tech. rep.) Burlington, VT: Polhemus Navigation Sciences, Inc.

DiVerdi, S., & Höllerer, T. (2007). Groundcam: A tracking modality for mobile mixed reality. *Proceedings of IEEE Virtual Reality Conference, 2007, VR '07,* 75–82.

Emura, S., & Tachi, S. (1994). Sensor fusion based measurement of human head motion. *3rd IEEE International Workshop on Robot and Human Communication (RO-MAN 94),* 124–129.

Foxlin, E. (1996). Intertial head-tracker sensor fusion by a complementary separate-bias Kalman filter. *Proceedings of 1996 Virtual Reality Annual International Symposium, VRAIS '96,* 185–194.

Foxlin, E. (2002a). Generalized architecture for simultaneous localization, auto-calibration, and map-building. *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems,* Vol. 1, 527–533.

Foxlin, E. (2002b). *Motion tracking requirements and technol-*

ogies, (Chap. 30, pp. 163–210). Mahwah, NJ: Lawrence Erlbaum.

Foxlin, E., Harrington, M., & Pfeifer, G. (1998). Constellation: A wide-range wireless motion-tracking system for augmented reality and virtual set applications. In M. F. Cohen (Ed.), *SIGGRAPH '98, Computer Graphics, Annual Conference on Computer Graphics & Interactive Techniques,* 371–378.

Foxlin, E., & Naimark, L. (2003). VIS-Tracker: A wearable vision-inertial self-tracker. *VR '03, Proceedings of the IEEE Virtual Reality 2003,* 199.

Friedmann, M., Starner, T., & Pentland, A. (1992a). Device synchronization using an optimal linear filter. *SI3D '92, Proceedings of the 1992 Symposium on Interactive 3D Graphics,* 57–62.

Friedmann, M., Starner, T., & Pentland, A. (1992b). Synchronization in virtual realities. *Presence: Teleoperators and Virtual Environments, 1*(1), 139–144.

Gelb, A. (1974). *Applied optimal estimation.* Cambridge, MA: MIT Press.

Haas, M. W. (1983). Visually coupled systems as simulation devices. *Proceedings of AIAA Flight Simulation Technologies Conference.*

Haas, M. W. (1984). Visually coupled systems as simulation devices. *Journal of Aircraft, 21*(8), 639–640.

Holloway, R. L. (1995). *Registration errors in augmented reality systems.* Ph.D. thesis, University of North Carolina at Chapel Hill.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME—Journal of Basic Engineering, 82*(Series D), 35–45.

Kocian, D. F. (1977). A visually-coupled airborne systems simulator (VCASS)—An approach to visual simulation. *Proceedings of the 1977 Image Conference,* 14–23.

Liang, J., Shaw, C., & Green, M. (1991). On temporal-spatial realism in the virtual reality environment. *1991 Symposium on User Interface Software and Technology.*

List, U. H. (1983). *Nonlinear prediction of head movements for helmet-mounted displays.* Tech. Rep. AFHRL-TP-83-45, Williams Air Force Base, Air Force Human Resources Laboratory, Operations Training Division.

Luinge, H., & Veltink, P. (2005). Measuring orientation of human body segments using miniature gyroscopes and accelerometers. *Medical & Biological Engineering & Computing, 43*(2), 273–282.

Maybeck, P. S. (1979). *Stochastic models, estimation, and con-trol,* Vol. 141 of *Mathematics in Science and Engineering.* New York: Academic Press.

Mazuryk, T., & Gervautz, M. (1995). Two-step prediction and image deflection for exact head tracking in virtual environments. *Computer Graphics Forum—Proceedings of EUROGRAPHICS 95,* Vol. 14, 30–41.

Meyer, K., Applewhite, H., & Biocca, F. (1992). A survey of position trackers. *Presence: Teleoperators and Virtual Environments,* 1(2), 173–200.

Pausch, R. (1991). Virtual reality on five dollars a day. *CHI '91: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 265–270.

Raab, F. H., Blood, E. B., Steiner, T. O., & Jones, H. R. (1979). Magnetic position and orientation tracking system. *IEEE Transactions on Aerospace and Electronic Systems,* AES-15(5), 709–718.

Rebo, R. K. (1988). *A helmet-mounted virtual environment display system.* Master's thesis, Air Force Institute of Technology.

Rebo, R. K., & Amburn, P. (1989). Helmet-mounted virtual environment display system. In J. T. Carollo (Ed.), *Proceedings of SPIE* Vol. 1116 of *Helmet-Mounted Displays,* 80–84.

Shin, H. J., Lee, J., Shin, S. Y., & Gleicher, M. (2001). Computer puppetry: An importance-based approach. *ACM Transactions on Graphics, 20*(2), 67–94.

Smith, B. R., Jr. (1984). Digital head tracking and position prediction for helmet mounted visual display systems. *Proceedings of 22nd AIAA Aerospace Sciences Meeting.*

So, R. H. Y., & Griffin, M. J. (1992). Compensating lags in head-coupled displays using head position prediction and image deflection. *Journal of Aircraft, 29*(6), 1064–1068.

Sutherland, I. E. (1965). The ultimate display. *Proceedings of IFIP Congress,* Vol. 2, 506–508.

Sutherland, I. E. (1968). A head-mounted three dimensional display. *Proceedings of the 1968 Fall Joint Computer Conference, AFIPS,* Vol. 33, Part 1, 757–764.

Tak, S., Song, O.-Y., & Ko, H.-S. (2002). Spacetime sweeping: An interactive dynamic constraints solver. *IEEE Computer Society: CA'02, Proceedings of Computer Animation,* 261–270.

Vallidis, N. (2002). *WHISPER: A spread spectrum approach to occlusion in acoustic tracking.* Ph.D thesis, University of North Carolina at Chapel Hill.

Vlasic, D., Adelsberger, R., Vannucci, G., Barnwell, J., Gross, M., Matusik, W., et al. (2007). Practical motion capture in everyday surroundings. *ACM Transactions on Graphics, 26*(3), 35.

Wang, J.-F., Azuma, R. T., Bishop, G., Chi, V., Eyles, J., & Fuchs, H. (1990). Tracking a head-mounted display in a room-sized environment with head-mounted cameras. *SPIE 1990 Technical Symposium on Optical Engineering and Photonics in Aerospace Sensing,* Vol. 1290 of *Helmet-Mounted Displays II,* pp. 47–57.

Ward, M., Azuma, R. T., Bennett, R., Gottschalk, S., & Fuchs, H. (1992). A demonstrated optical tracker with scalable work area for head-mounted display systems. *Symposium on Interactive 3D Graphics,* 43–52.

Welch, G. (1995). *Hybrid self-tracker: An inertial/optical hybrid three-dimensional tracking system.* (Tech. rep. TR95-048). University of North Carolina at Chapel Hill, Department of Computer Science.

Welch, G., & Bishop, G. (1995). *An introduction to the Kalman filter.* (Tech. rep. TR95-041). University of North Carolina at Chapel Hill, Department of Computer Science.

Welch, G., & Bishop, G. (1997). SCAAT: Incremental tracking with incomplete information. In T. Whitted (Ed.), *SIGGRAPH '97, Computer Graphics, Annual Conference on Computer Graphics & Interactive Techniques,* 333–344.

Welch, G., & Bishop, G. (2008). The Kalman filter web site. Available at http://www.cs.unc.edu/~welch/kalman/.

Welch, G., Bishop, G., Vicci, L., Brumback, S., Keller, K., & Colucci, D. (2001). High-performance wide-area optical tracking: The HiBall tracking system. *Presence: Teleoperators and Virtual Environments, 10*(1), 1–21.

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

Wikipedia. (2008). Edwin Albert Link. Available at http://en.wikipedia.org/wiki/Edwin_Albert_Link.

Woltring, H. J. (1974). New possibilities for human motion studies by real-time light spot position measurement. *Biotelemetry, 1*(3), 132–146.

This article has been cited by:

1. Sandru Florin, Nanu Sorin, Negrea Romeo, Popescu Anca, Fericel Sergiu. Data Acquisition and Processing with Fusion Sensors, Used in Algorithms for Increasing Accuracy of Position Measurement of Mobile Robots 565-578. [Crossref]

2. Ahmet Kose, Aleksei Tepljakov, Sergei Astapov. Real-time localization and visualization of a sound source for virtual reality applications 1-6. [Crossref]

3. Jinwei Gu, Xiaodong Yang, Shalini De Mello, Jan Kautz. Dynamic Facial Analysis: From Bayesian Filtering to Recurrent Neural Network 1531-1540. [Crossref]

4. Chi Zhang, Amirhossein Taghvaei, Prashant G. Mehta. Attitude estimation of a wearable motion sensor 4570-4575. [Crossref]

5. Ahmed Nejmedine Machraoui, Mohamed Fethi Diouani, Jamila Ghrab, Mounir Sayadi. Automated detection and tracking of phlebotominaes 206-211. [Crossref]

6. Chern Hong Lim, Ekta Vats, Chee Seng Chan. 2015. Fuzzy human motion analysis: A review. *Pattern Recognition* **48**:5, 1773-1796. [Crossref]

7. Cody O. Karutz, Jeremy N. Bailenson. Immersive Virtual Environments and the Classrooms of Tomorrow 290-310. [Crossref]

8. Yang Zheng, Ka-Chun Chan, Charlie C. L. Wang. Pedalvatar: An IMU-based real-time body motion capture system using foot rooted kinematic model 4130-4135. [Crossref]

9. Chengkai Xiang, Hui Huang Hsu, Wu Yuin Hwang, Jianhua Ma. Comparing Real-Time Human Motion Capture System Using Inertial Sensors with Microsoft Kinect 53-58. [Crossref]

10. Alberico Menozzi, Brian Clipp, Eric Wenger, Jared Heinly, Enrique Dunn, Herman Towles, Jan-Michael Frahm, Gregory Welch. Development of vision-aided navigation for a wearable outdoor augmented reality system 460-472. [Crossref]

11. Matjaž Mihelj, Domen Novak, Samo Begus. Tracking the User and Environment 53-95. [Crossref]

12. Ovgu Ozturk, Tomoaki Matsunami, Yasuhiro Suzuki, Toshihiko Yamasaki, Kiyoharu Aizawa. 2012. Real-time tracking of humans and visualization of their future footsteps in public indoor environments. *Multimedia Tools and Applications* **59**:1, 65-88. [Crossref]

13. Ajoy Kumar Das, Achintya Haldar, Subrata Chakraborty. 2012. Health Assessment of Large Two Dimensional Structures Using Limited Information: Recent Advances. *Advances in Civil Engineering* **2012**, 1-16. [Crossref]

14. Angelo Maria Sabatini. 2011. Estimating Three-Dimensional Orientation of Human Body Parts by Inertial/Magnetic Sensing. *Sensors* **11**:12, 1489-1525. [Crossref]

15. Angelo Maria Sabatini. 2011. Kalman-Filter-Based Orientation Determination Using Inertial/Magnetic Sensors: Observability Analysis and Performance Evaluation. *Sensors* **11**:10, 9182-9206. [Crossref]

16. Tim Wrigley. Motion Sensors in Osteoarthritis 183-220. [Crossref]

17. Hao Yang, Juntao Ye. A calibration process for tracking upper limb motion with inertial sensors 618-623. [Crossref]

18. Qiong Wu, Pierre Boulanger. Real-Time Estimation of Missing Markers for Reconstruction of Human Motion 161-168. [Crossref]

19. Jerry van den Heuvel, Jacco Bikker. Detecting 3D Position and Orientation of a Wii Remote Using Webcams 448-458. [Crossref]

20. Marco T. Angulo, Jaime Moreno, Ricardo Lazaro. Robust dissipative observer design for nonlinear systems 111-115. [Crossref]

21. Jesse Fox, Dylan Arena, Jeremy N. Bailenson. 2009. Virtual Reality. *Journal of Media Psychology* **21**:3, 95-113. [Crossref]