Motion Tracking: No Silver Bullet, but a Respectable Arsenal

Greg Welch
University of North Carolina at Chapel Hill

Eric Foxlin
InterSense

This article introduces the physical principles underlying the variety of approaches to motion tracking. Although no single technology will work for all purposes, certain methods work quite well for specific applications.

If you read the surveys of motion tracking systems,[1-5] one thing that will immediately strike you is the number of technologies and approaches—a bewildering array of systems operating on entirely different physical principles, exhibiting different performance characteristics, and designed for different purposes. So why does the world need so many different tracking products and research projects to do essentially the same thing?

Just as Brooks argued in his famous article on software engineering[6] that there is no single technique likely to improve software engineering productivity an order of magnitude in a decade, we'll attempt to show why no one tracking technique is likely to emerge to solve the problems of every technology and application.

But this isn't an article of doom and gloom. We'll introduce you to some elegant trackers designed for specific applications, explain the arsenal of physical principles used in trackers, get you started on your way to understanding the other articles in this special issue, and perhaps put you on track to choose the type of system you need for your own computer graphics application. We hope this article will be accessible and interesting to experts and novices alike.

What is motion tracking?

If you work with computer graphics—or watch television, play video games, or go to the movies—you are sure to have seen effects produced using motion tracking. Computer graphics systems use motion trackers for five primary purposes:

- *View control.* Motion trackers can provide position and orientation control of a virtual camera for rendering computer graphics in a head-mounted display (HMD) or on a projection screen. In immersive systems, head trackers provide view control to make the computer graphics scenery simulate a first-person viewpoint, but animations or other nonimmersive applications might use handheld trackers.
- *Navigation.* Tracked devices help a user navigate through a computer graphics virtual world. The user might point a tracked wand to fly in a particular direction; sensors could detect walking-in-place motion for virtual strolling.
- *Object selection or manipulation.* Tracked handheld devices let users grab physical surrogates for virtual objects and manipulate them intuitively. Tracked gloves, acting as virtual surrogates for a user's hands, let the user manipulate virtual objects directly.
- *Instrument tracking.* Tracked tools and instruments let you match virtual computer graphics representations with their physical counterparts—for example, for computer-aided surgery or mechanical assembly.
- *Avatar animation.* Perhaps the most conspicuous and familiar use of trackers has been for generating realistically moving animated characters through full-body motion capture (MoCap) on human actors, animals, and even cars.

No silver bullet

Our experience is that even when presented with motion tracking systems that offer relatively impressive performance under some circumstances, users often long for a system that overcomes the shortcomings related to their particular circumstances. Typical desires are reduced infrastructure, improved robustness, and reduced latency (see the sidebar, "Tracking Latency"). The only thing that would satisfy everyone is a magical device we might call a "tracker-on-a-chip." This ToC would be all of the following:

- *Tiny*—the size of an 8-pin DIP (dual in-line package) or even a transistor;
- *Self-contained*—with no other parts to be mounted in the environment or on the user;

0272-1716/02/$17.00 © 2002 IEEE

Tracking Latency

Have you seen those so-called "gourmet" cookie stands in convenience stores or fast-food restaurants? They usually include a sign that boasts "Made fresh daily!" Unfortunately, while cookie baking might indeed take place daily, the signs don't actually give you the date on which the specific cookies being sold were baked!

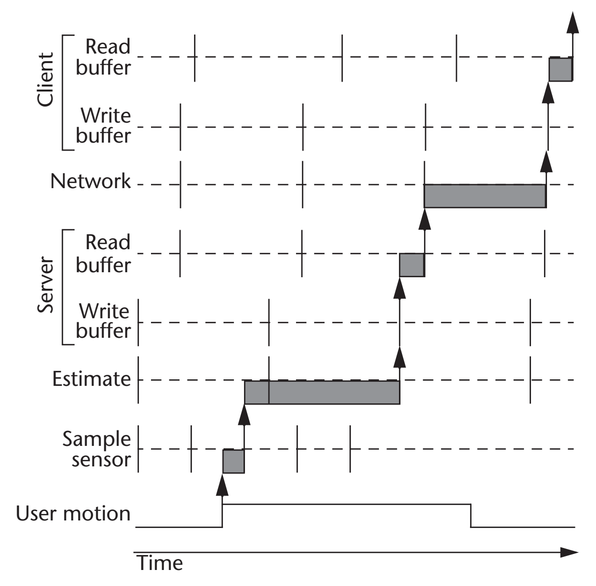
We've found a related common misperception about delay or latency in interactive computer graphics in general, and in tracking in particular. While the inverse of the estimate rate (the period of the estimates) contributes to the latency, it doesn't tell the entire story. Consider our imaginary tracker-on-a-chip. If you send its 1,000-Hz estimates halfway around the world over the Internet, they will arrive at a rate of 1,000 Hz, but quite some time later.

Similarly, within a tracking system, a person moves, the sensors are sampled at some rate, some computation is done on each sample, and eventually estimates pop out of the tracker. To get the entire story, you must consider not only the rate of estimates, but also the length of the pipeline through which the sensor measurements and subsequent pose estimates travel.

As Figure A illustrates, throughout the pipeline there are both fixed latencies, associated with well-defined tasks such as sampling the sensors and executing a function to estimate the pose, and variable latencies, associated with buffer operations, network transfers, and synchronization between well-defined but asynchronous tasks. The variable latencies introduce what's called latency jitter.

Here again there's no silver bullet. In 1995 Azuma showed that motion prediction can help considerably, to a point.[1,2] The most basic approach is to estimate or measure the pose derivatives and to use them to extrapolate forward from the most recent estimate—which is already old by the time you get to see it—to the present time. The problem is that it's difficult to predict what the user will choose (has chosen) to do very far in the future.

Azuma pointed out that the task is like trying to drive a car by looking only in the rear-view mirror. The driver must predict where the road will go, based solely on the view of the past and knowledge of roads in general. The difficulty of this task depends on how fast the car is going and on the shape of the road. If the road is straight and remains so, the task is easy. If the road twists and turns unpredictably, the task is impossible.



A **Typical tracker pipeline.**

References

1. R. Azuma, *Predictive Tracking for Augmented Reality*, PhD dissertation, tech. report TR95-007, Univ. North Carolina, Chapel Hill, Dept. Computer Science, 1995.
2. R. Azuma and G. Bishop, "A Frequency-Domain Analysis of Head-Motion Prediction," *Proc. Ann. Conf. Computer Graphics and Interactive Techniques* (Proc. Siggraph 95), ACM Press, New York, 1995, pp. 401-408.

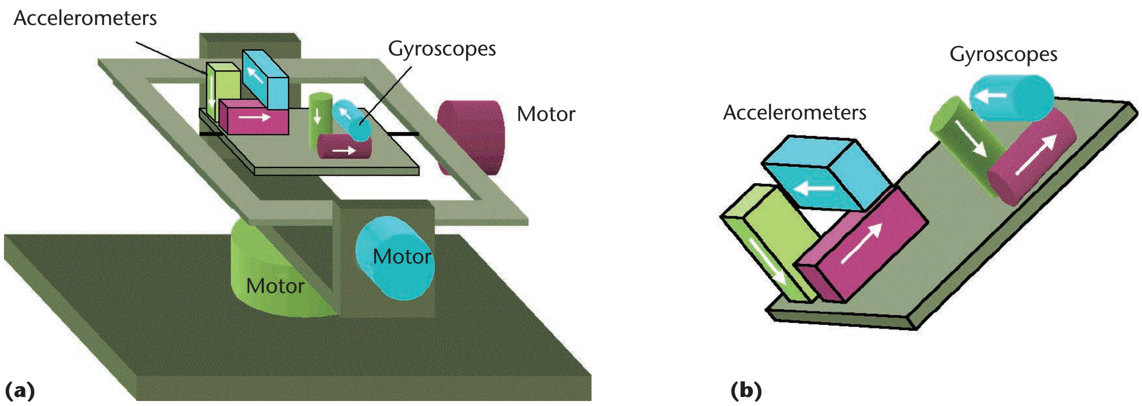
- *Complete*—tracking all six degrees of freedom (position and orientation);
- *Accurate*—with resolution better than 1 mm in position and 0.1 degree in orientation;
- *Fast*—running at 1,000 Hz with latency less than 1 ms, no matter how many ToCs are deployed;
- *Immune to occlusions*—needing no clear line of sight to anything else;
- *Robust*—resisting performance degradation from light, sound, heat, magnetic fields, radio waves, and other ToCs in the environment;
- *Tenacious*—tracking its target no matter how far or fast it goes;
- *Wireless*—running without wires for three years on a coin-size battery; and
- *Cheap*—costing $1 each in quantity.

If this magic ToC existed, we would use it for everything. The reality is that every tracker today falls short on at least seven of these 10 characteristics, and that number is unlikely to shrink much in the foreseeable future.

But all is not lost! Researchers and developers have pragmatically and cleverly exploited every available physical principle to achieve impressive results for specific applications. We'll start with an overview of some of the available ammunition and the strengths and weaknesses of each and then look at some specific applications and the tracking technologies that have been employed successfully in each.

Available ammunition

Although designers have many pose estimation algorithms to choose among, they have relatively few sensing technologies at their disposal. In general, the technologies sense and interpret electromagnetic fields or waves, acoustic waves, or physical forces. Specifically, motion tracking systems most often derive pose estimates from electrical measurements of mechanical,



1 (a) Stable-platform (gimbaled) INS. (b) Strapdown INS.

inertial, acoustic, magnetic, optical, and radio frequency sensors.

Each approach has advantages and limitations. The limitations include modality-specific limitations related to the physical medium, measurement-specific limitations imposed by the devices and associated signal-processing electronics, and circumstantial limitations that arise in a specific application. For example, electromagnetic energy decreases with distance, analog-to-digital converters have limited resolution and accuracy, and body-worn components must be as small and lightweight as possible. Although alternative classifications are possible, we discuss the available ammunition using a traditional medium-based classification.

Mechanical sensing

Arguably the simplest approach conceptually, mechanical sensing typically involves some form of a direct physical linkage between the target and the environment. The typical approach involves an articulated series of two or more rigid mechanical pieces interconnected with electromechanical transducers such as potentiometers or shaft encoders. As the target moves, the articulated series changes shape and the transducers move accordingly. Using a priori knowledge about the rigid mechanical pieces and online measurements of the transducers, you can estimate the target's position (one end of the link) with respect to the environment (the opposite end).

This approach can provide very precise and accurate pose estimates for a single target, but only over a relatively small range of motion—typically one cubic meter. In his pioneering HMD work in 1968, Sutherland built a mechanical tracker composed of a telescoping section with a universal joint at either end. While Sutherland and his colleagues found the system too cumbersome in practice, they relied on it as a "sure method" of determining head pose. The most common uses of mechanical sensing today are for boom-type tracked displays that use counterweights to balance the load and for precision 3D digitization over a small area. Commercial examples include the Boom 3C by FakeSpace and the FaroArm by Faro Technologies.

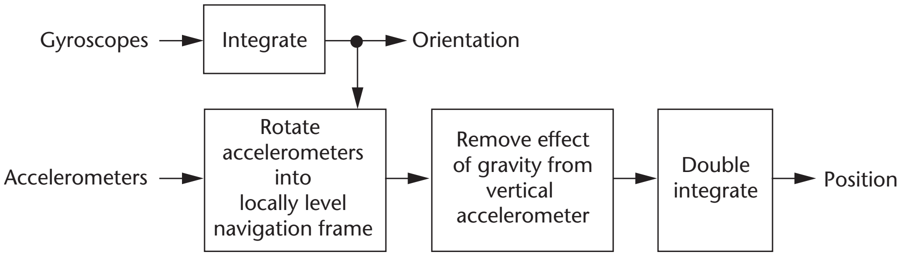
Articulated haptic devices such as the Phantom by SensAble Technologies inherently include mechanical tracking of the force-feedback tip. These devices need to know the tip position to apply appropriate forces, and the electromechanical devices typically used to provide the forces can also be used to sense the position.

Inertial sensing

Inertial navigation systems (INSs) became widespread for ships, submarines, and airplanes in the 1950s, before virtual reality or computer graphics were even conceived, but they were the last of the six ammunition technologies to be introduced for computer graphics input devices. The reason is straightforward: an INS contains gyroscopes, and early high-accuracy spinning-wheel gyroscopes weighed far too much to be attached to a person's body. Not until the advent of MEMS (microelectronic mechanical systems) inertial sensors in the 1990s did the development of inertial input devices begin.

Originally, inertial navigation systems were built with a gimbaled platform (see Figure 1a) stabilized to a particular navigation reference frame (such as north-east-down) by using gyroscopes on the platform to drive the gimbal motors in a feedback loop. The platform-mounted accelerometers could then be individually double-integrated to obtain position updating in each direction, after compensating for the effect of gravity on the vertical accelerometer. Most recent systems are of a different type, called strapdown INS (see Figure 1b), which eliminates mechanical gimbals and measures a craft's orientation by integrating three orthogonal angular-rate gyroscopes strapped down to the craft's frame. To get position, three linear accelerometers, also affixed to the moving body, measure the acceleration vector in body-frame, which is then rotated into navigation coordinates using the current rotation matrix as determined by the gyroscopes. The result is a navigation-frame acceleration triad just like that measured by the accelerometers in the stable-platform INS, which can be gravity-compensated and double-integrated in the same way. Figure 2 illustrates this flow of information.

Inertial trackers might appear to be the closest thing to a silver bullet of all the ammunition technologies we describe here. Gyroscopes and accelerometers are already available in chip form, and within the next decade we expect to see a single-chip six-axis strapdown inertial measurement unit—that is, with three gyroscopes and three accelerometers. Inertial sensors are completely self-contained, so they have no line-of-sight



2 **Basic strap-down inertial navigation algorithm.**

Tracking Performance Specifications and Requirements

In deciding the quality of tracking required for an application involving visual simulation such as virtual reality, there are several possible goals:

- The user feels presence in the virtual world.
- Fixed virtual objects appear stationary, even during head motion (perceptual stability).
- No simulator sickness occurs.
- Tracking artifacts don't affect task performance.
- Tracking artifacts remain below the detection threshold of a user looking for them.

Several types of tracking errors can contribute in varying degrees to destroying the sense of presence or perceptual stability, causing sickness, or degrading task performance. Various authors and manufacturers have focused on different specifications or defined them differently, and every type of tracker has its own complicated idiosyncrasies that would require a thick document to characterize in complete detail. However, Table A presents six specifications that can capture the essential aspects of tracking performance that affect human perception of a virtual environment while a tracked object is still (static) or moving (dynamic).

There's no clearly defined distinction between spatial jitter and creep, as they could be thought of representing the high- and low-frequency portions of a continuous noise spectrum. A reasonable cutoff might be to designate as creep any motion slower than a minute hand in orientation (0.1 degree per second) and slower than 1 mm per second in translation, with everything else called jitter.

Table A. Tracking performance specifications.

Static	**Dynamic**
Spatial distortion. Repeatable errors at different poses in the working volume, including effects of all sensor scale factors, misalignments, and nonlinearity calibration residuals, and repeatable environmental distortions.	*Latency.* The mean time delay after a motion until corresponding data is transmitted. It's possible to specify the latency of the tracker and other subsystems separately, but they don't simply add up.
Spatial jitter. Noise in the tracker output that causes the perception of the image shaking when the tracker is actually still.	*Latency jitter.* Any cycle-to-cycle variations in the latency. When moving, this will cause stepping, twitching, multiple image formation, or spatial jitter along the direction the image is moving.
Stability or creep. Slow but steady changes in tracker output may appear over time. The cause might be temperature drift or repeatability errors if the tracker is power-cycled or moved and returned to the same pose.	*Dynamic error (other than latency)*. This error type includes any inaccuracies that occur during tracker motion that can't be accounted for by latency or static inaccuracy (creep and spatial distortion). This might include overshoots generated by prediction algorithms or any additional sensor error sources that are excited by motion.

requirements, no emitters to install, and no sensitivity to interfering electromagnetic fields or ambient noise. They also have very low latency (typically a couple of milliseconds or less), can be measured at relatively high rates (thousands of samples per second), and measured velocity and acceleration can generally be used to predict the pose of a head or a hand 40 or 50 ms into the future. Good inertial sensors also offer extremely low jitter (see the sidebar, "Tracking Performance Specifications and Requirements").

The weakness that prevents inertial trackers from being a silver bullet is *drift*. If one of the accelerometers has a bias error of just 1 milli-g, the reported position output would diverge from the true position with an acceleration of 0.0098 m/s^2. After a mere 30 seconds, the estimates would have drifted by 4.5 meters! If you look closely at Figure 2, you can see that an orientation error of 1 milliradian coming from the gyroscopes would produce a gravity compensation error of 1 milli-g on one of the horizontal accelerometers, causing just this calamity.

Even very good gyroscopes (the kind you wouldn't want to wear on your head) drift by a milliradian within a short time. Nevertheless, given the advantages we've enumerated, inertial sensors can prove very valuable when combined with one or more other sensing technologies, such as those we describe next. Inertial sensors have provided the basis for several successful hybrid systems.

Acoustic sensing

Acoustic systems use the transmission and sensing of sound waves. All known commercial acoustic ranging systems operate by timing the flight duration of a brief ultrasonic pulse.

In contrast, in 1968 Sutherland built a continuous carrier-phase acoustic tracking system to supplement his mechanical system.[7] This system used a continuous-wave source and determined range by measuring the phase shift between the transmitted signal and the signal detected at a microphone. Meyer and colleagues point out that this "phase-coherent" method enables continuous measurement without latency but can only measure relative distance changes within a cycle.[3] To measure absolute distance, you need to know the starting distance and then keep track of the number of accumulated cycles. Another problem, which could be the reason no successful implementation of the phase-coherent approach has been developed, is the effect of multipath reflections. *Multipath,* a term also associated with radio transmission, indicates that the signal received is often the sum of the direct path signal and one or more reflected signals of longer path lengths. Because walls and objects in a room are extremely reflective of acoustic signals, the amplitude and phase of the signal received from a continuous-wave acoustic emitter in a room will vary drastically and unpredictably with changes in the receiver's position.

An outstanding feature of pulsed time-of-flight acoustic systems is that you can overcome most multipath reflection problems by waiting until the first pulse arrives, which is guaranteed to have arrived via the direct path unless the signal is blocked. The reason this method works for acoustic systems but not for radio frequency and optical systems is that sound travels relatively slowly, allowing a significant time difference between the arrival of the direct path pulse and the first reflection.

Point-to-point ranging for unconstrained 3D tracking applications requires transducers that are as omnidirectional as possible, so that the signal can be detected no matter how the emitter is positioned or oriented in the tracking volume. To achieve a wide beam width, you must use small speakers and microphones with active surfaces a few millimeters in diameter. This is convenient for integration into human motion tracking devices and helps reduce off-axis ranging errors, but the efficiency of an acoustic transducer is proportional to the active surface area, so these small devices can't offer as much range as larger ones.

To improve the range, most systems use highly resonant transducers and drive them with a train of electrical cycles right at the resonant frequency to achieve high amplitude. This results in a received waveform that "rings up" gradually for about 10 cycles to a peak amplitude then gradually rings down. For a typical envelope-peak detection circuit, this means the point of detection is delayed about 10 cycles—about 90 mm—from the beginning of the waveform. By detecting on the second or third cycle instead of the 10th, you can greatly reduce the risk of multipath reflection.

In our experience, this is one of the most important issues for accurate ultrasonic tracking outside of controlled laboratory settings, and it is the crux of how InterSense's ultrasonic ranging technology remains accurate at longer ranges than others.

The physics of ultrasonic waves in air and transducer design dictate other design trade-offs and considerations as well. Most ambient noise sources fall off rapidly with increasing frequency, so operating at a higher frequency is beneficial for avoiding interference, and the shorter wavelengths offer higher resolution. However, selecting a higher frequency reduces the range because of problems with transducer size and frequency-dependent attenuation of sound in air, which starts to play a significant role by 40 kHz and becomes the dominant factor in limiting range by 80 kHz, depending on humidity.

Ultrasonic trackers typically offer a larger range than mechanical trackers, but they're not a silver bullet. Their accuracy can be affected by wind (in outdoor environments) and uncertainty in the speed of sound, which depends significantly on temperature, humidity, and air currents. A rule of thumb is that the speed of sound changes about 0.1 percent per degree Fahrenheit of temperature differential. This corresponds to about a one-millimeter error per degree Fahrenheit at one meter.

Acoustic systems' update rate is limited by reverberation. Depending on room acoustics and tracking volume, it may be necessary for the system to wait anywhere from 5 to 100 ms to allow echoes from the previous measurement to die out before initiating a new one, resulting in update rates as slow as 10 Hz. The latency to complete a given acoustic position measurement is the time for the sound to travel from the emitter to the receivers, or about one millisecond per foot of range. This is unaffected by room reverberation and is usually well under 15 ms in the worst case. However, in a purely acoustic system with a slow update rate, the need to wait for the next measurement also affects system latency.

Acoustic systems require a line of sight between the emitters and the receivers, but they're somewhat more tolerant of occlusions than optical trackers (which we discuss later) because sound can find its way through and around obstacles more easily. Finally, we have yet to see a purely acoustic tracker that doesn't go berserk when you jingle your keys.

You can address most of the shortcomings we've mentioned by building a hybrid system that combines acoustic sensors with others that have complementary characteristics—inertial sensors, for example.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**

fastcase®
Smarter legal research.