Tracking for Virtual Environments

David E. Johnson

I. INTRODUCTION

It might be argued that tracking is a fundamental discriminator for virtual environments in comparison to general interactive programs. While display technologies take a variety of forms, such as HMD’s, CAVE’s, workbenches, and fishtank displays, all depend on head tracking to retain the illusion of immersion. Furthermore, intuitive interactivity with a computer environment is also severely curtailed without tracking of a user’s arms and hands. This chapter will explore the mathematical background and technologies for effective tracking of a person’s head and hands. Finger-based tracing through gloves will be explored in later material on haptic and tactile feedback.

Tracking is essentially the recovery of position through measurements to known landmarks, often referred to as markers or sometimes beacons. Some of the earliest techniques for tracking were developed for navigation, astronomy, and mapping using triangulation. Tracking also draws upon estimation theory to merge multiple measurements into a single, more accurate, estimation of position.

Current tracking systems are based on optical, electromagnetic, mechanical, and inertial measurements; each has unique advantages and disadvantages. Before discussing these systems, it is worthwhile to provide some metrics to evaluate the performance to tracking systems (from The Science of Virtual Reality and Virtual Environments, Roy Kalawsky):

- **Static and dynamic accuracy**: The positional and rotational accuracy of a tracker is of critical importance. Accuracy may vary for markers that are still compared to moving markers. In general, it is difficult to characterize a tracker’s accuracy without full disclosure of the test environment - for example, ultrasound trackers are very sensitive to temperature variations...even a vent blowing air can negatively impact accuracy. Rotational accuracy can be tightly tied to positional as rotation is often inferred from the position of a few closely spaced markers.
- **Latency**: The latency of the computation is the time from initiating a measurement to when the marker position is available to the user. Some tracking systems send multiple signals to recover position, so the overall latency is increased. Furthermore, estimation of position from raw analog signals may require a fast DSP to minimize computational latency. Finally, tracking systems often have a dedicated base station that must communicate with the computer. Early systems using slow serial lines incurred significant communication latency.
- **Update rate**: Update rate is often incorrectly favored over latency as a measure of system performance. A tracker may be able to process 1,000 samples per second, but if there is an (exaggerated) 500 ms latency, then the samples are mostly useless. The update rate is often bound by the sensing technology, such as camera refresh, or signal decay.
- **Registration**: Registration refers to a tracker’s ability to maintain a consistent zero. For translation, minor fluctuations in “home” are easily adapted to, whereas a tilted horizon in a HMD can be disconcerting.
- **Drift**: Drift is related to registration, but can be more difficult to deal with. While a device may record fairly accurate \( \delta X \) changes in position, the error accumulates over time. This is often a result of integrating accelerations to get positions, or could even be an effect from a source “warming up” over time. In any case, having a hand drift away from a body can be disconcerting for a user.
- **Signal to noise ratio and jitter**: The signal to noise ratio impacts accuracy, but also the overall quality of tracking. A system may be able to very accurately position a marker, but only after extensive filtering to remove noise. Such filtering introduces latency into the system. A less noisy tracker may be preferable to a more accurate but noisy one.
- **Degrees of Freedom (DOF)**: The degrees of freedom typically is either 3DOF, for positional tracking, or 6DOF for positional and rotational tracking. Some systems create a 6DOF system out of 3DOF tracking by tracking the corners of a small reference frame, but rotational accuracy can suffer.

Furthermore, there are several other subjective criteria which can play as important a role in choosing the appropriate tracking technology:
Fig. 1. A position on a map can be found by measuring the bearings of two landmarks, drawing lines through the landmarks oriented at the bearing, and finding the point of intersection of the two lines.

- **Encumbrance:** Many trackers require the markers to be connected by wire, which can limit ease of use. Most current trackers have a wireless option. Weight and marker size can be an issue; for example, the UNC HiBall tracker works well on HMD’s but is inappropriate for hand tracking. Even visual markers for optical systems can be brushed off or gotten dirty, making them difficult for some tasks.

- **Range:** Most trackers in the 1990’s had fairly limited range of just a few feet. With the growth in importance of motion caption (MoCap), wider range trackers have been developed. Wireless technology has aided this direction.

- **Distortion:** A tracker may be susceptible to distortions from the environment. Metal induces distortions to electromagnetic trackers and enclosed spaces may confuse ultrasonic trackers.

- **Line-of-sight requirements:** Optical trackers are sensitive to line-of-sight problems, where a marker can be occluded by the environment or the user’s body. Multiple cameras can alleviate these problems, but at additional cost.

- **Number of markers tracked:** Some systems can only track a few markers, possibly by reducing the update rate and time-slicing the process, while others can track several hundred with only a computational cost. MoCap applications are especially demanding of high numbers of markers.

### A. Triangulation Basics

Triangulation methods were among the earliest used for tracking. Given a map and a compass, triangulation recovers position by determining the bearing to two landmarks. Lines from each landmark in the bearing direction intersect at a unique point which is the estimated position. Figure 1 shows a hiker recognizing two mountain peaks, measuring the angle $A$ (relative to north) to one peak and the angle $B$ to the other peak. Drawing those lines on the map through the landmarks locates the hiker.

Triangulation also refers to making bearing measurements to an object from two different locations a known distance apart. In this case, the law of sines is used to recover the position of the object.

First, let us review some standard notion and properties for triangles. For a triangle with vertices $A$, $B$, and $C$, the sides opposite the vertices are labeled $a$, $b$, and $c$ (Figure 2(a)). Somewhat confusingly, the angle associated with each vertex is also labeled $A$, $B$, and $C$, respectively. This is because a triangle has six degrees of freedom, and given three side lengths, the angles uniquely identify the triangle. In the following discussion, $A_{xy}$ will be used when necessary to specify the coordinates of the vertex rather than the angle.

Additionally, recall that for a right triangle, when $B$ is the right angle, the length of $a$ is $\sin(A)b$ and the length of $c$ is $\cos(A)b$. 
Fig. 2. (a) The vertices and internal angles of a triangle are labeled $A$, $B$, and $C$, while the lengths of the opposing sides are $a$, $b$, and $c$. (b) The position of a boat can be inferred by measuring the angle to the boat from two different lighthouses a known distance apart.

**Question:** If the hypotenuse of a right triangle has length 1, then what famous Theorem does the following represent?

$$\sin^2(A) + \cos^2(A) = 1,$$

As an example problem (Figure 2(b)), if two lighthouses, lighthouse $A$ and lighthouse $B$, are one kilometer apart and simultaneously measure their bearings $A$ and $B$ to a boat $C$, then the length $a$ or $b$ may be recovered from the law of sines

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}. \quad (1)$$

Since the internal angles of a triangle sum to $180^\circ$, angle $C = 180^\circ - A - B$. Length $c$ is also known, it is the distance between the two lighthouses. Therefore, if $A = 30^\circ$ and $B = 50^\circ$, then

$$b = \frac{(1\,km)(\sin 50^\circ)}{\sin(180^\circ - 30^\circ - 50^\circ)},$$

yielding a length $b \approx 0.78$. The equivalent computation can be done using the measured $A$ instead. The projection of the ship onto the baseline forms two right triangles and the position of the ship is then

$$C_{xy} = A_{xy} + b \sin(A)\vec{y} + b \cos(A)\vec{x} = B_{xy} + a \sin(B)\vec{y} - a \cos(B)\vec{x},$$

where $\vec{x}$ and $\vec{y}$ are the axes of the coordinate system defined by $\overrightarrow{AB}$ and the line between the ship and that baseline, respectively.

**B. Applications to Optical Tracking**

Optical tracking systems are attractive due to their flexibility, high-performance, accuracy, and latency. Some disadvantages are line-of-sight issues and relatively poor rotational tracking. Optical systems are most often used in motion capture; the need to have multiple cameras inside the VR workspace limits their application in VE’s like CAVEs. The basic mathematical approach to optical tracking uses triangulation.

Optical tracker systems place markers, either reflective or active LED, on the tracked person. Multiple cameras at fixed locations “see” the markers and infer the bearing to the marker based on the camera parameters. This is accomplished by measuring the disparity in location of the marker on each camera’s focal plane, $d_L$ for the left camera and $d_R$ for the right. Figure 3 shows the setup for cameras with parallel view directions. In this case, the
Fig. 3. The disparity in projection of an object onto two cameras’ shared focal plane can be used to find the depth of the object from the cameras. The left image shows a 2D example for depth, while the right image shows the full general case.

Depth $Z$ is obtained by comparing the height and width of two similar triangles. The first similar set is for the left side

$$\frac{Z}{B_L} = \frac{f}{d_L}$$

and the second for the right side

$$\frac{Z}{B_R} = \frac{f}{-d_R},$$

where $B_L + B_R = B$. Since neither $B_L$ nor $B_R$ are known, as the projection of the object on $B$ is not measured by the cameras, $Z$ cannot be found directly from either equation. Instead, the two equations can be rearranged and combined as follows to find $Z$.

$$B_L = \frac{Zd_L}{f}$$

$$B_R = \frac{-Zd_L}{f}$$

$$B_L + B_R = \frac{Z(d_L - d_R)}{f}$$

$$\frac{Z}{B} = \frac{f}{(d_L - d_R)}$$

$$Z = \frac{fB}{(d_L - d_R)}$$

The quantity $(d_L - d_R)$ is called the stereo disparity. The peculiar use of $-d_R$ in the left figure is because the coordinate system for left and right cameras should be similarly oriented, as shown in the right figure. Seeing the similar triangles is easier when $d_R$ is negative. Also, the leftward facing x-axis in the right figure is because the image plane is often considered in front of the center of focus of the lens and the axis flips when the image plane crosses the plane of the lens.
C. Placement in 3D

In 3D, similar arguments are used to find the full X,Y,Z position of the object in space. Assuming the world coordinate system lies on the lens center of the left camera, then

\[
\begin{align*}
Z &= \frac{fB}{x_L - x_R} \\
X &= \frac{x_LZ}{f} \\
Y &= \frac{y_LZ}{f}
\end{align*}
\]

The formulas for X is often given for an origin halfway between the two cameras. In this case,

\[
\begin{align*}
X &= \frac{x_LZ}{f} - \frac{B}{2} \\
X &= \frac{x_LfB}{f(x_L - x_R)} - \frac{B}{2} \\
X &= \frac{2xLB}{2(x_L - x_R)} - \frac{B(x_L - x_R)}{2(x_L - x_R)} \\
X &= \frac{B(x_L + x_R)}{2(x_L - x_R)}
\end{align*}
\]

which is the form it is often shown in.

D. Optical tracker resolution exercise

The ability of an optical tracker to resolve the position of an object is highly dependent on the resolution available to measure the stereo disparity. Imagine building a tracker from cheap webcams. Webcams have roughly 1000 x 1000 pixels measured on a sensor 10mm x 10mm. Further pretend the focal length is 10mm. Take two of these webcams and fixture them 100mm apart. If the object is 1000mm away, centered above the left camera, then the disparity will be

\[
\begin{align*}
x_L &= \frac{10mm \times 1000mm}{1000mm} = 1000pixels, \\
x_R &= \frac{10mm \times 1000mm}{10mm} = 101pixels, \\
(x_L - x_R) &= -100pixels.
\end{align*}
\]

The shift to the neighboring 101th pixel happens when

\[
\begin{align*}
(x_L - x_R) &= -101pixels, \\
(0 - x_R) &= -101pixels. \\
x_R &= \frac{10mm \times 1000mm}{101pixels} = 1000pixels, \\
z &= \frac{10mm \times 1000mm}{101pixels} = 990.1mm.
\end{align*}
\]

Thus, at that position, this simple tracker can only resolve changes in Z of around 10mm. Note that this is not constant, and part of the engineering of an optical tracker must be deciding where to use your pixels for highest sensitivity.

**Questions:** What happens when the object starts 100mm away? What if the baseline is increased to 1000mm?
II. ACoustIC TRACKERS

Triangulation is not the only means of determining the position of an object by making measurements. Acoustic trackers determine the distance from an emitter to a sensor and compute position by combining measurements from multiple locations.

Just as the math underlying optical tracking draws upon simple trigonometric concepts, the math underlying acoustic tracking uses basic geometric concepts. The basic approach uses an emitter that ‘chirps’, creating a sound wave that propagates outward until a sensor detects it. The time between emitting and receiving is the time-of-flight for that signal. Distance is then computed by multiplying the time-of-flight by the velocity of sound

$$D = \Delta t \times v_{\text{sound}}$$

Sadly, this simple equation is already fraught with hidden complication. First, while the speed of sound is fortunately slow enough to allow the $\Delta t$ to be measured (think of the difficulty of doing this approach with light!), it is also slow enough to create unavoidable latency in the system. Sound travels roughly 1 ft/ms, so even a small space creates several milliseconds of latency per measurement, and several are needed for position measurements.

Second, the speed of sound varies with temperature and pressure. An acoustic tracker should be calibrated for use in Salt Lake City, UT vs Duluth, MN, for example. Furthermore, if a portion of the tracked space is being blown on by a heating vent, the speed of sound in its volume changes by

$$\Delta c = 167.6 \frac{m}{sec} + 0.6 \frac{m}{(sec)(degree)} \Delta T_{\text{Kelvin}}.$$  

Assuming a temperature change of $10^\circ K$ from the vent, then the speed of sound and the associated distance estimate changes by around 4%, or 40mm per meter of distance.

A. Position from Distance

One sound source and one sensor determines their relative separation, $D_1$. In effect, the sensor is constrained to lie on a sphere of radius $D_1$ centered on the source. A second source in a different location finds a different distance, $D_2$, and the source is constrained to lie on a sphere around that source. The intersection of the two spheres is a circle; points on that circle satisfy both constraints. A third source creates yet another constraint sphere. The intersection of all these spheres yields two points, either of which could be the position of the sensor. One of these points is typically on the “wrong” side of the sources and can be eliminated. These concepts are illustrated in Figure 4.
Fig. 5. The wire generates a magnetic field along its length. The coiled wire field reinforces itself into a strong magnetic dipole.

B. Ultrasonic trackers

A typical configuration for an acoustical tracker uses emitters at 40kHz, which is in the ultrasonic range. Three emitters are used with three sensors to compute a local reference frame - this yields both position and orientation. Each emitter is activated in sequence and the distances to the three sources measured in parallel. A delay must be introduced in between each emitter’s activation to allow echoes to attenuate.

**Question:** What is the impact of a moving the sensor array while measurements are being taken?

Currently, ultrasonic trackers are either used for low-end applications like shutter glasses on regular monitors, or as part of a hybrid tracking system.

C. Phase Coherence Ultrasonic Trackers

It is possible to get more accuracy and less latency using acoustic signaling. Instead of measuring the arrival of a wavefront at a sensor, the phase of the signal is measured. For a known frequency signal and the change in phase between emission and reception, a distance proportional to the phase shift is computed as

\[
\delta = \frac{c}{f} \times \frac{\phi}{2\pi}
\]

where \(c\) is the speed of sound, \(f\) is the frequency of the signal, and \(\phi\) is the measured phase shift. At 40 kHz, a measured phase shift of \(\pi/4\) corresponds to 1.03mm. The main difficulty with this approach is that a phase shift of \(\pi/4\) is indistinguishable to shifts of \(2\pi + \pi/4\). Since phase can be measured continuously, it is possible to continually track the phase shift and maintain knowledge of relative position.

In 1966, the MIT Lincoln wand used phase coherent tracking for position and orientation of a pointer. Sutherland experimented with a similar tracker for his HMD in 1968. This approach seems interesting and it is not clear why it does not have greater prominence. Signal to noise due to echoes may be a problem.

III. Electromagnetic Trackers

Both optical and acoustic trackers are susceptible to line-of-sight problems. While this is less of an issue with trackers for HMD’s, where the marker can be mounted on top of the user’s head, occlusions are a serious problem for hand tracking. Electromagnetic trackers do not suffer from occlusion problems and were the dominant form of tracking until the late 1990’s.

A. Physics Background

A magnetic field is the relativistic by-product of a moving electric charge. A stationary electric charge has a spherically symmetric electric field. When this charge moves, the field is distorted, causing a magnetic field. The magnetic field is a vector field.
Fig. 6. The induced current is proportional to the amount of magnetic flux passing through the sensor loop. (a) The strongest induced current is when the sensor loop is perpendicular to the magnetic field. (b) As the loop rotates, the cross-sectional area of the loop aligned with the field reduces by the cosine of the angle. (c) Each coil in the base station generates a changes magnetic field in sequence. The sensor coils simultaneously read these field and the induced current gives orientation and distance. The z-axis field is not shown.

Inductance is the proportionality between the electric current and magnetic flux - that is, an electric current generates a magnetic field.

\[ I = \frac{\Phi}{i} \]

For a solenoid, or coil of wire, the magnetic flux is multiplied by the number of turns, \( N \) in the coil

\[ I = N \frac{\Phi}{i} \]

Figure 5 shows how a wire with an electrical current generates a field and by bending the wire, the field is added into a unified magnetic dipole.

Just as an electric current creates a magnetic field, the above equations also show that a changing magnetic field induces an electric current. This somewhat magical effect is the basis for electromagnetic trackers.

An alternating current (AC) is applied to a coil in the tracker base station, generating a changing magnetic field. The magnetic field passes through a coil in the sensor, inducing a current, which can be measured. The strength of the induced current varies with the cube of the distance to the base station as well as the cosine of the angle between the sensor coil and magnetic field.

Look at Figure 6, which shows how the flux passing through the sensor coil varies as the sensor is tilted. Figure 6(c) illustrates the operation of an AC electromagnetic tracker. The base station passes an AC current through each coil of the base station in sequence. These fields induce in turn currents in all three sensor coils – the strength of these currents for all the fields gives an estimate of orientation and distance.

Magnetic fields are not easily blocked, so electromagnetic trackers do not suffer from line-of-sight issues that acoustic and optical ones do. However, they do suffer from field distortions. Just as the source magnetic field induces a current in the source coil, they also induce current in any metal object in the vicinity. This induced current also generates its own magnetic field, which induces its own current in the sensor coil. Typically three orthogonal fields are induced in sequence, each one read by three orthogonal sensors, allowing recovery of distance and orientation information.

The has led to the amusing situation that many high-tech VR installations are built primarily of low-tech wood. However, most large building are built with steel girders spaced every 20 feet or so, and these are difficult to avoid. More recently, trackers have used direct current to generate a static magnetic field. This field can be measured by Hall effect sensors, which measure magnetic field strength perpendicular to a sensor plate.

The DC trackers have their own problems to overcome. Three orthogonal magnetic fields are generated in sequence and measured by three magnetometers in the sensor. As each field is turned off and on, the field changes, inducing currents and field creation in metal objects, similar to AC systems. The sensor must wait for these “eddy currents” to dissipate before making measurements. Furthermore, the Earth has a magnetic field of its own, as does
the building, so after generating the three orthogonal fields, the base station is turned off and this ambient magnetic
field is measured. The ambient field can then be subtracted from the next set of measurements.

These requirements limit the update rate for electromagnetic trackers to around one hundred Hz, with accuracies
of a 0.1 inch and 0.5°. However, this is superior to acoustic trackers. Additionally, the base station and sensors can
be arranged more flexibly than optical trackers. Until the development of inertial trackers, these systems dominated
VR installations.

IV. INERTIAL TRACKERS

While other technologies provide adequate performance for position tracking, head-tracking for HMD’s has more
stringent requirements.

*Question: Why is head-tracking for a HMD harder than head tracking in a CAVE environment?*

Inertial trackers are currently the highest performance solution for low-encumbrance, low-latency position and
orientation tracking. They are based on accelerometers and gyroscopes to directly compute motion.

An accelerometer is essentially a damped mass-spring system. An external force will displace a mass on the
spring, the displacement is proportional to the acceleration. Angular acceleration is measured by the displacement
of a gyroscope. In MEMS systems, a small sense element distorts under acceleration, producing a differential
capacitance or piezoelectric voltage or other effect.

These trackers are called *inertial* because they rely on inertia as described by Newton’s First Law of Motion,
which states that “a body at rest stays at rest and a body in motion stays in motion unless disturbed by an unbalanced
force”. If these forces can be measured accurately enough, then the motion of the object can be determined as well.

Using position as an example, if an initial position, \( x(0) \), and velocity, \( v(0) \), are given, then the expected position
over time is

\[
x(t) = x(0) + v(0)t.
\]

In the presence of a constant force, for example gravity \( g \), the velocity changes continuously

\[
v(t) = v(0) + \frac{g}{m}t,
\]

where \( m \) is the mass of the object. These equations are better recognized as being integrals, where change in velocity
is the integral of acceleration over time and change in position is the integral of velocity over time. Mathematically,

\[
\Delta v(t) = \int_0^T a(t) \, dt
\]

\[
\Delta x(t) = \int_0^T v(t) \, dt.
\]

By measuring the acceleration with an accelerometer, the integral for velocity is approximated at small time
steps, and position can be numerically integrated as well. This technology was originally developed for spacecraft
and missiles; MEMS technology has made it small enough to use worn on a head or hand. Car air-bags are common
application of MEMS accelerometers; more recent uses are in laptops for hard drive head parking and in shoes for
MP3 player integration.

*Questions: What happens if your time-step is too large? What impact does round-off error have on the tracking?
What should the relative sensitivity of positional and angular acceleration be?*

A pure inertial tracker is open loop - that is, there is no additional feedback for error correction. Thus, the tracked
position drifts away from the true position over time and this error grows without bound. For this reason, inertial
trackers are almost always combined with other tracking approaches which provide absolute measurements. The
combination of different measurements is part of estimation theory and filtering, which is addressed in the next
section.

V. MECHANICAL TRACKERS

Mechanical trackers typically have the best accuracy, lowest latency, and smallest jitter. However, they have
the most limited range and are perhaps the most encumbering. For some applications, they should be seriously
considered, but they are not flexible enough for most VR applications.
VI. ESTIMATION AND FILTERING

Thus far, we have assumed that every measurement is perfect and straight-forward application of the relevant math (triangulation, time-of-flight, or motion integrals) provides the correct answer. However, we have already hinted at reasons measurements may be less than perfect: temperature distortions for acoustic tracers, bearing and baseline measurement errors for triangulation, and finite resolution for optical trackers and accelerometers.

Uncertainty in measurements may be propagated through the mathematics to estimate uncertainty in position measurements. For example, if the time of flight measurement on an acoustic tracker is only precise to $\pm 0.1 \text{ msec}$, then the distance from time of flight equation,

$$D = \Delta t \times v_{\text{sound}},$$

would provide a $D$ precise to $\pm (0.0001 \text{ sec}) (340 \frac{m}{\text{sec}})$ or $\pm 0.034 m$ when using a $v_{\text{sound}}$ of $340 \frac{m}{\text{sec}}$. If temperature changes affect the speed of sound, then further uncertainty is propagated.

Recall that for acoustic tracking, multiple distances were used to constrain the position of the sensor through the intersection of spheres. Using error bounds, the solution is no longer constrained to a point, instead a region of space potentially satisfies the constraint equation. Figure 7 shows a two-dimensional visualization of such a region. Note that the gray regions that satisfy the constraints are not simple boxes, thus the position cannot be correctly specified as a point plus or minus some value.

A. Statistics Primer

In reality, uncertainty and noise in measurements often follow a normal distribution, whereas the distribution implied by plus or minus some value is a uniform distribution. For discrete events, the distribution is described by a set of events, each with an associated probability. Distributions follow three axioms:

1. For any set $\mathcal{E} \in \mathcal{F}$, where $\mathcal{F}$ is the set of all possible events for the random variable,

$$P(\mathcal{E}) \geq 0.$$

2. The probability of some event in the allowable set happening is

$$P(\mathcal{F}) = 1.$$

3. For $n$ disjoint sets $\mathcal{E}_1, \mathcal{E}_2 \ldots \mathcal{E}_n$

$$P(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \cdots) = \sum_{i=1}^{n} P(\mathcal{E}_i)$$

For probabilities on a continuous domain, a probability distribution is a function mapping intervals to probabilities for which the three axioms are true. Why map intervals? Look at the case for height, a continuous function. Yet the probability of a person being exactly two meters tall is essentially zero, it makes more sense to discuss the
probability of a person being $6m \pm 0.01m$. This mapping is through the probability density function, $f(x)$, which maps real numbers to a probability value, the distribution is the integral of the density function over some interval.

$$P(a \leq X \leq b) = \int_a^b f(x) \, dx.$$ 

This probability density function describes the distribution. A uniform distribution has a probability density function $\frac{1}{b - a}$. A normal distribution has a density function which is the familiar bell-shaped curve. A normal distribution is also called a Gaussian distribution and looks like

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x-\mu)^2 / 2\sigma^2}.$$ 

There are two undescribed variables in this function. The mean of the function, $\mu$, is the average. Most people are familiar with the formula for the mean of a discrete set of $N$ values,

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

How can this be applied to a continuous distribution? Well, for a discrete set of events, some may be more probable than others, such that the above equation may be reinterpreted as

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \ldots + f_k x_k}{N}$$

where $k$ is the number of different observed $x$ values and $\frac{f_i}{N}$ is the normalized discrete probability function as $N$ goes to infinity. This gives the connection to a continuous probability function, where the mean, $\mu$, is

$$\mu = \int x f(x) \, dx.$$ 

The mean is also the expected value, $E[X]$, of a random value $X$.

The second undefined variable, $\sigma^2$, is the variance of a distribution. Variance is the spread of a distribution – two distributions can have the same mean, but very different variances. For a discrete distribution, the variance is

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2$$

which is also $var(X) = E((X - \bar{x})^2)$. The standard deviation is more familiar to most people and is just the square root of the variance, $\sigma$. For a continuous distribution, the variance is

$$\sigma^2 = \int (x - \mu)^2 f(x) \, dx.$$ 

The mean and variance are the first and second moments of a distribution. There are other moments, but a normal distribution is completely described by mean and variance and often denoted $X \sim N(\mu, \sigma^2)$.

B. Noise

Variation, or uncertainty, in measurement is a result of noise in the measurement. The error from noise often has a Gaussian distribution. This is a result of The Central Limit Theorem, which states that the sum of a sequence of random finite-variance variables will approach a normal distribution as the number of variables increases. This is independent of the distributions of the independent variables (see http://www.stat.sc.edu/west/javahtml/CLT.html for an example of rolling dice and the central limit theorem).

For the purposes of tracking, error in measurement is likely the cumulative effect of multiple small measurements, such as start time and end time for multiple emitters and sensors. Characterizing the error in distance and position
measurements as Gaussian distributions is then fairly reasonable, even without intense analysis of the component error distributions.

Noise with a non-zero mean has a bias. Bias must be dealt with by other means, but noise with a zero mean can be dealt with by filtering. Filtering removes the impact of noise on the measurement by providing an accurate estimate of the true state.

Let us examine the time-of-flight equations again while using the statistical terms reviewed above.

\[ D = \Delta t \times v_{\text{sound}}, \]

First we looked at it using perfect measurements, then with an implied uniform error distribution of \( \pm \) some error. If, instead, the time is a normal distribution with \( \mu = t \) and variance \( \sigma^2 \), or \( \mathcal{X} \sim N(t, \sigma^2) \) then scaling it by \( v_{\text{sound}} \) results in a normal distribution \( \mathcal{X} \sim N(\mu v_{\text{sound}}, v_{\text{sound}}^2 \sigma^2) \), as seen by scaling the integral form of variance above.

Assume a standard deviation in time measurement of \( 0.1 \) m/sec, or a variance of \( 0.01 \) m/sec\(^2\). Then the variance in distance measurement will be \( (0.01 \text{msec}^2)((0.34 \frac{m}{\text{msec}})^2, \) or \( 0.0011m \). This is a standard deviation of \( 34mm \), which matches our intuition from having done the \( \pm \) computation earlier. But now, the measurement is fully characterized by a Gaussian distribution, and we can use this to merge multiple measurements into a single, better estimate of the true distance.

C. Filtering

Given a set of measurements with Gaussian noise characteristics, how can the noise be reduced? A method to reduce noise is called a filter.

A simple scenario for a filter is to recover the true state of some object given repeated noisy measurements. The state simply might be an object’s position, or it might encode any number of variables, such as position, orientation, velocity, etc.

If the noise is Gaussian, then over time, the histogram of samples approximates the Gaussian describing the noise, with its mean at the true value of the state. This is true by definition, really, but it is somewhat non-intuitive. Another way to look at it is that the random noise tends to cancel itself out over time, leaving true value. So in some ways, filtering is not so difficult!

However, a related problem to that given above is using repeated measurements to recover the state of an object while the state evolves. In this case, the mean is a moving target, so the previous approach of taking a large number of samples to fill in the Gaussian doesn’t work. A common approach in this instance is to use a moving average filter, which averages the last \( n \) measurements to estimate the current state. The mean at time \( k \) is then

\[ \bar{x}_k = \frac{1}{n} \sum_{i=k-n+1}^{k} x_i \]

Note that this is more efficiently computed as

\[ \bar{x}_k = \bar{x}_{k-1} + \frac{1}{n}(x_k - x_{k-n}). \]

The main problem with this approach is lag. The estimated state lags the true state. Assume you are trying to measure a tracker moving at a constant speed along the x-axis. If there is little noise, when \( n = 5 \), the estimated state lags the true state by two time steps. For objects undergoing acceleration, the estimate becomes even worse.

The exponentially-weighted moving filter tries to address this by weighting more recent measurements more heavily. Updating the mean estimate then becomes

\[ \bar{x}_k = \alpha \bar{x}_{k-1} + (1 - \alpha)x_k, \]

where \( \alpha \) is typically \( \frac{n}{n+1} \), but other weightings are possible. While this does reduce the lag in the state estimate, it also tends to emphasize the impact of noise as well.

When filters are looked at from a Fourier point of view, filters are basically low-pass filters. Noise is mostly made up of high-frequency components, so by removing that frequency response, noise is reduced. However, the natural evolution of the state may include high-frequency components and removing those reduces the ability of the estimate to respond to those changes.
One direction for improvement to these basic filters is to include an internal system model. If the desired state to be measured is position, but measurements on velocity are available, then it would make sense to somehow combine the direct measurements of position with estimates of position based on the equations of motion. Potentially the velocity measurements could be very accurate while the position measurements very poor (such as is the case for hybrid inertial-acoustic trackers). In this case, it makes sense to weight the combination of the two towards the velocity measurement.

D. Kalman Filter

The Kalman filter is a filtering algorithm that does just that. It uses maintains a state vector, a linear state evolution matrix, and estimates of reliability to provide optimal combination of measurements in estimating the true state.

1) Covariance Matrix: First, one more definition. The Kalman filter applies to states that are multi-dimensional vectors. Estimates of these states are Gaussians, but multi-dimensional Gaussians. The mean, \( \mu \), is then a vector. Each dimension also has a variance, but instead of a vector of variances, a matrix of covariances is used. The diagonal of the covariance matrix contains the variances, but in general, an element of the matrix \( \text{cov}(X, Y) \) is

\[
\text{cov}(X, Y) = \sigma_{XY} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}).
\]

The covariance matrix gives a measure of the correlation between dimensions in the samples. A covariance matrix with only diagonal non-zeros elements corresponds to a multi-dimensional Gaussian aligned along the axes. A generic covariance matrix would appear to be rotated.

2) Estimation: Combining an a priori estimate of state with a new measurement is the basic operation of Kalman filters. A slightly simpler case is the combination of two simultaneous measurements. Imagine taking a measurement of tracker position using an acoustic system. Because it is a relatively noisy tracker, the measurement \( x_a \) has a relatively large uncertainty, which is encoded as a Gaussian \( X_a \sim N(x_a, \sigma_a^2) \). At the same time, a high-resolution optical tracker measures its position, \( X_o \sim N(x_o, \sigma_o^2) \). The uncertainty for the optical tracker is much smaller than that of the acoustic tracker.

Is the acoustic tracker measurement worthless? On some level it seems that it would “corrupt” the optical measurement. However, as long as the estimate of their uncertainty is accurate, they can be combined to provide an even more accurate measurement with smaller overall uncertainty. The derivation is based on a least squares minimization; we give the result for a combined estimate \( X_{\text{combined}} \sim N(\mu_{\text{combined}}, \sigma_{\text{combined}}^2) \)

\[
\mu_{\text{combined}} = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_o^2} x_a + \frac{\sigma_o^2}{\sigma_a^2 + \sigma_o^2} x_o
\]

\[
\sigma_{\text{combined}}^2 = \frac{\sigma_a^2 \sigma_o^2 (\sigma_a^2 + \sigma_o^2)}{(\sigma_a^2 + \sigma_o^2)^2}
\]

The Kalman gain, \( K \), is a term used to weight the different estimates, and is simply a rewriting of the terms from the combined estimate above.

\[
\mu_{\text{combined}} = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_o^2} x_a + \frac{\sigma_o^2}{\sigma_a^2 + \sigma_o^2} x_o
\]

\[
\mu_{\text{combined}} = x_a + \frac{\sigma_a^2}{\sigma_a^2 + \sigma_o^2} [x_o - x_a]
\]

\[
K = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_o^2}
\]

Think about what happens when combining measurements with the same variance or ones with very different variances. Do the equations meet your expectations?

The Kalman filter uses the same approach, except that one measurement is a real measurement and the other “measurement” is an estimate of the state and uncertainty of that estimate based on prior measurements and an internal state evolution model. The Kalman filter also maintains a covariance matrix for the internal state, which lets it decide how heavily to weight the prediction versus the new measurement.