LECTURE 10: TRACKING

Ehsan Aryafar
earyafar@pdx.edu
http://web.cecs.pdx.edu/~aryafare/VR.html

Recall: Motion of a Point in a 3D World

- Consider motion of a point (x,y,z) in 3D, then the coordinate values would be a function of time:

$$(x(t), y(t), z(t))$$

- Velocity and acceleration can be represented with respect to the three axis

$$(v_x, v_y, v_z) \qquad (a_x, a_y, a_z)$$

- The magnitude of velocity can be calculated as

$$\sqrt{v_x^2 + v_y^2 + v_z^2}$$

Recall: 3D Angular Velocity and Acceleration

- Consider a rigid-body rotating around a unit vector v
- This rotation can be decomposed into rotation around x, y, and z axis, resulting in the following 3D angular velocity

$$(\omega_x, \omega_y, \omega_z)$$

- Angular velocity may also change over time, resulting in angular acceleration of pitch, yaw, and roll angles

$$(\alpha_x, \alpha_y, \alpha_z)$$

Do not mix linear velocity/acceleration with angular velocity/acceleration!

Recall: Numerical Simulation of Rigid-Bodies

- Virtual world contains many rigid-bodes
- Physics engine is a key component of VWG
- Job of the VWG is to determine the position and orientation of each and every object for any given time
  - This snapshot of the virtual world is also referred to as world state
  - With known state, we can use the chain of transformation to place objects onto the user display
- Typically, a rigid-body has six degrees of freedom (DoF)
- Sometime, DoF are lost due to constraints

Recall: Partial Differential Equations

Extra!
Not in Exam

- We model motion through PDEs
- Let x = (x_1, …, x_n) denote an n-dimensional state vector
  - x_i corresponds to a position or orientation for a rigid-body
  - Let the time derivative or velocity for each parameter be $\dot{x}_i = \frac{dx_i}{dt}$
  - To obtain the state at time t, the velocities need to be integrated over time

$$x_i(t) = x_i(0) + \int_0^t \dot{x}_i(s)ds$$

  - Thus, we can capture motion by solving the above integration

Recall: How to Derive the Integral Values?

Extra!
Not in Exam

- Time is discretized into steps
  - Delta-t is the step size or sampling rate
  - For example, sampling rate of 1 msec means the physics engine updates all parameters (positions and orientations) every 1 msec
  - Think of it as the frame rate for the physics engine
  - Physics engine frame rate is much higher than renderer and display frame rates to have a much higher accuracy
- Euler approximation for state variable x_i:

$$x_i((k+1)\Delta t) \approx x_i(0) + \sum_{j=1}^{k} \dot{x}_i(j\Delta t)\Delta t,$$

$$x_i[k+1] \approx x_i[k] + \dot{x}_i(k\Delta t)\Delta t$$

This Lecture: Tracking

- Tracking of the physical world is a key part of any VR system
  - Keep track of the motion/placement of real objects
- Highly accurate tracking has been enabled by commodity hardware components such as inertial measurement units (IMUs) and cameras
  - Have plummeted in size due to the smartphone industry

Three Types of Tracking in VR

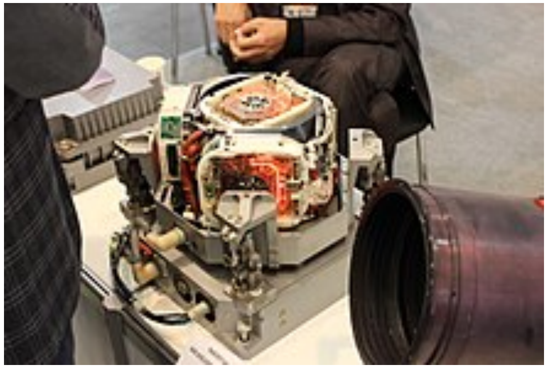
- The user's sense organ
  - If a display is attached to the sense organ, then position and orientation of the organ need to be tracked
  - Imagine wearing a VR headset: head tracking should happen
  - The visual system may need further tracking of eye as well
- The user's other body parts
  - When the user wants to see her body in the virtual world
  - For example, an interactive VR experience may need tracking of the facial expressions or hand gestures
- The rest of the environment
  - Physical objects in the physical world may be tracked
  - VR system may bring them into the virtual world for safety concerns
    - or to provide touch feedback for the user
  - In telepresence, live capture can bring physical world into virtual

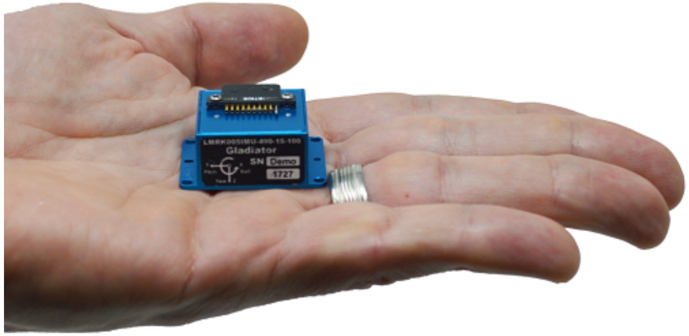
Outline

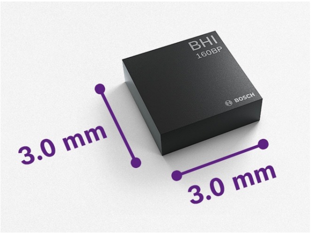
- Tracking orientation in a 2D world
  - Not covered: tracking 3DoF orientation of a 3D rigid body
- Tracking position and orientation (pose) together
  - In most systems requires line-of-sight (LoS) visibility between a fixed part of the physical world and the object being tracked
- In depth overview of several practical tracking solutions

Inertial Measurement Unit

- An inertial measurement unit (IMU) works by detecting linear acceleration using one or more accelerometers and rotational rate (angular velocity) using one or more gyroscopes. Some also include a magnetometer which is commonly used as a heading reference. Typical configurations contain one accelerometer, gyro, and magnetometer per axis for each of the three principal axes: pitch, roll and yaw.
- Applications: orientation of anything with attached IMU







IMUs have gone from large mechanical systems to cheap, microscopic MEMS circuits

Tracking 2D Orientation

- Imagine we mount a gyroscope to a merry-go-round
  - Measure the angular velocity
- In practice, the measured and true velocity are different
- Consider the following simple model

Estimated angular velocity → $\hat{\omega}$; offset → a; scale → b; True Angular Velocity → ω

$$\hat{\omega} = a + b\,\omega$$

$$\hat{\omega} - \omega = a + b\,\omega - \omega = a + \omega(b - 1)$$

The biggest contributor to the error above is the calibration error!

Drift Error

- Let d(t) denote a function of time called the drift error

Actual Orientation — Estimated Orientation

$$d(t) = \theta(t) - \hat{\theta}(t)$$

- Suppose orientation at time 0 is 0 and the angular velocity is constant. Then drift error would be

$$d(t) = (\hat{\omega} - \omega)t = (a + b\,\omega - \omega)t = (a + \omega(b - 1))t$$

- Tracking error increases at a faster rate as the head rotates more quickly

Four Problems to Solve For Effective Tracking

- Calibration: Assume access to a high quality sensor and a few low quality sensors. The output of the low quality sensors could be calibrated to mimic the high quality sensor.
- Integration: Sensors output their measurement in discrete times, so their outputs need to be integrated over time
  - Recall Euler integration
- Registration: The initial orientation must be determined through an additional sensor or a startup procedure
- Drift Error: As the error grows over time, other sensory data may be needed to compensate for it

Calibration

- Suppose we have access to a high quality sensor that reports angular velocity with high accuracy
- The goal of calibration is to have a mapping that maps the output of cheaper sensors to the output of good sensor
- **Methodology:**
  - Attach all sensors to a movable surface and rotate them all with the same speed
  - Obtain *n* samples of angular measurements for both the high quality and low quality sensors
  - Calibration: calibrate the output of each low quality sensor to the output of the high quality one by solving an optimization problem

Calibration Optimization

- Minimize the sum of square errors across n samples

Low quality sensor output → $$\sum_{i=1}^{n}(\hat{\omega}_i - \hat{\omega}'_i)^2$$ ← High quality sensor output

- We can select constants c1 and c2 that optimize the error

$$\sum_{i=1}^{n}(c_1 + c_2\,\hat{\omega}_i - \hat{\omega}'_i)^2$$

- This is a classical regression problem referred to as linear least square (easy to solve)
- The calibrated output is derived from raw output as

From here on, we only work with the calibrated output

$$\hat{\omega}_{cal} = c_1 + c_2\,\hat{\omega}$$

Integration

- Sensor outputs arrive at discrete intervals
  - Example: Oculus gyroscope provides a measurement every 1 ms
- The orientation at time t can be estimated by integration as

Initial Orientation

$$\hat{\theta}[k] = \theta(0) + \sum_{i=1}^{k} \hat{\omega}[i]\Delta t$$

Calibrated angular velocity at time i*Delta_t

- Quite often, the above equation is written in incremental form

$$\hat{\theta}[k] = \hat{\omega}[k]\Delta t + \hat{\theta}[k-1]$$

- The angular velocity could be the velocity at the beginning of the interval, end of the interval, or an average

Registration

- In the previous slide, we assumed initial orientation is known
- Registration problem: the initial alignment between the real and virtual worlds
- Possible solutions
  - Set the initial alignment to zero
    - Essentially as soon as the system is booted or as soon as the user puts on the headset (if there is an in-head sensor)
      - Handing the headset to another person in the middle can be problematic
  - Have a fixed point in the physical world that corresponds to the zero alignment
    - Example: if the VR headset is facing the monitor
    - This solution would require a sensor that can measure orientation w.r.t. the physical world. For example, with the Oculus Rift, the user faces a stationary camera, which corresponds to the forward direction.

Drift Error Correction

- If the gyroscope was perfectly calibrated, drift would grow over time, because of
  - Quantized output values, sampling rate limitations, noise
- Solution: Have an independent additional sensor
  - Merge the two readings
  - Example: Suppose there is a camera that takes pictures of merry-go-round and derives an orientation estimate
- A classical approach, is a complementary filter, which interpolates the two readings

Gain parameter

$$\hat{\theta}_c[k] = \alpha\hat{\theta}_d[k] + (1 - \alpha)\hat{\theta}[k]$$

Instantaneous measurement from the secondary sensor

Estimated orientation by integrating the gyroscope

Drift Error Continued

- The gain parameter is typically a very small value
  - Causes gradual impact of the instantaneous estimate
  - The secondary sensor typically has a much slower sampling rate
    - Use the most recent sample across many estimate derivations
- Using simple linear algebra, the drift error correction through complementary filter can be re-written as

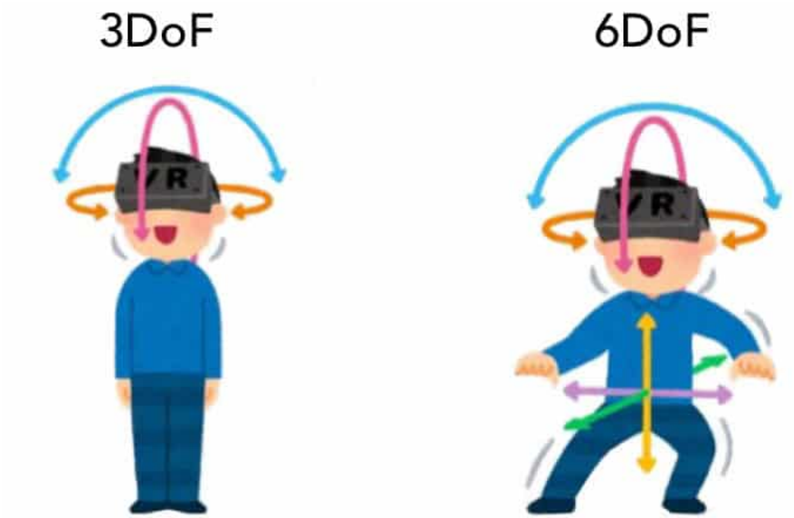
$$\hat{\theta}_c[k] = \hat{\theta}[k] - \alpha\hat{d}[k]$$

$$\hat{d}[k] = \hat{\theta}_d[k] - \hat{\theta}[k]$$

An estimate of the drift error at stage k

Recall: 3DoF vs 6DoF

- Mobile VR headsets like Oculus Go, Samsung Gear VR, and Google Daydream only have rotational tracking (3DoF)
  - Can look up/down/sides, but cannot laterally move in the VR world
  - 3DoF controllers are similar (act as laser selection pointers)
  - IMUs are used to track 3 DoF for orientation
- High end VR systems feature positional tracking (6DoF)
  - You can move in the virtual world by walking
  - With 6DoF controllers, you can interact with virtual objects by moving your hands in the real world
  - Most important use case: head tracking
  - 6 DoF enables forming T_{eye} from sensory data



Tracking Position

- IMU also provides three linear acceleration components through accelerometer
  - If we integrate twice, that would give the position
  - Drift error in gyroscope is result of one-time integration
  - If we use accelerometer for position, the drift error would be a result of two times of integration over accelerometer
    - Would result in very high inaccuracy even after 1 second
    - Would also be of little use if velocity is constant (no acceleration)

In commercial systems, twice integration over acceleration is indeed used! We will talk about this very soon!

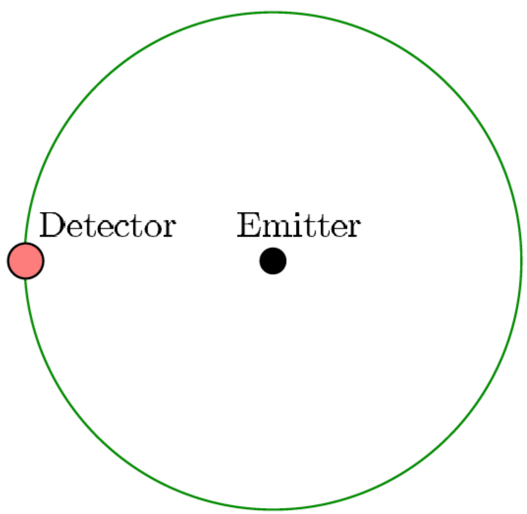
In the next three slides we discuss an abstract solution to solve position tracking!

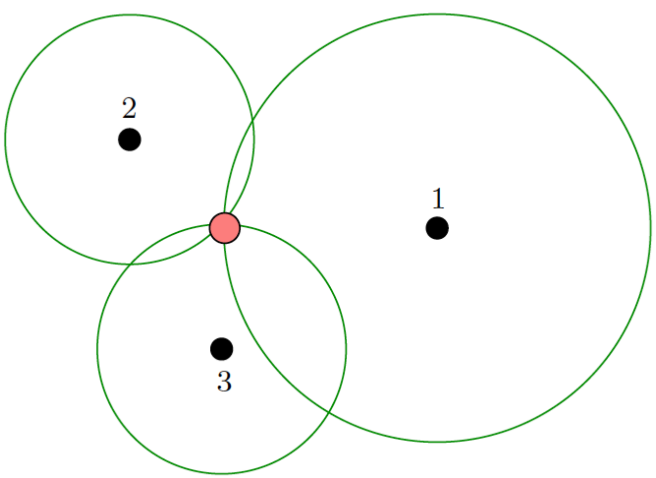
Transmitting Custom Waves

- Use of IMU for tracking is passive
- An active approach can be used for pose detection
  - Light and sounds are bad wave choices
  - Possible choices: Ultrasound, electromagnetic, and infrared
- Consider transmitting an ultrasound pulse from a speaker (emitter) and a microphone as the receiver (detector)
  - Our ears cannot hear them
  - Above 20,000 Hz
- Assume emitter and detector are perfectly synchronized
  - Time of arrival/flight (TOA) can be calculated
  - Speed is known (330 m/sec for ultrasound)
    - Distance can be estimated
    - Position of detector can be narrowed down to a sphere

Trilateration in 2D Case

- With two transmitters, position of the detector could be narrowed down to two points
- With three or more emitters, position can be uniquely identified
- Also, roles can be reversed
    - Object being tracked become emitter
    - Several fixed receivers are placed around it
- The method of combining these measurement is called **trilateration**





Use of EWs Instead of Ultrasound

- Electromagnetic waves such as light, radio, or infrared can be used instead of ultrasound
  - Would be impossible to directly measure the propagation time
- However, trilateration can still be used
  - Distance can be estimated based on power degradation

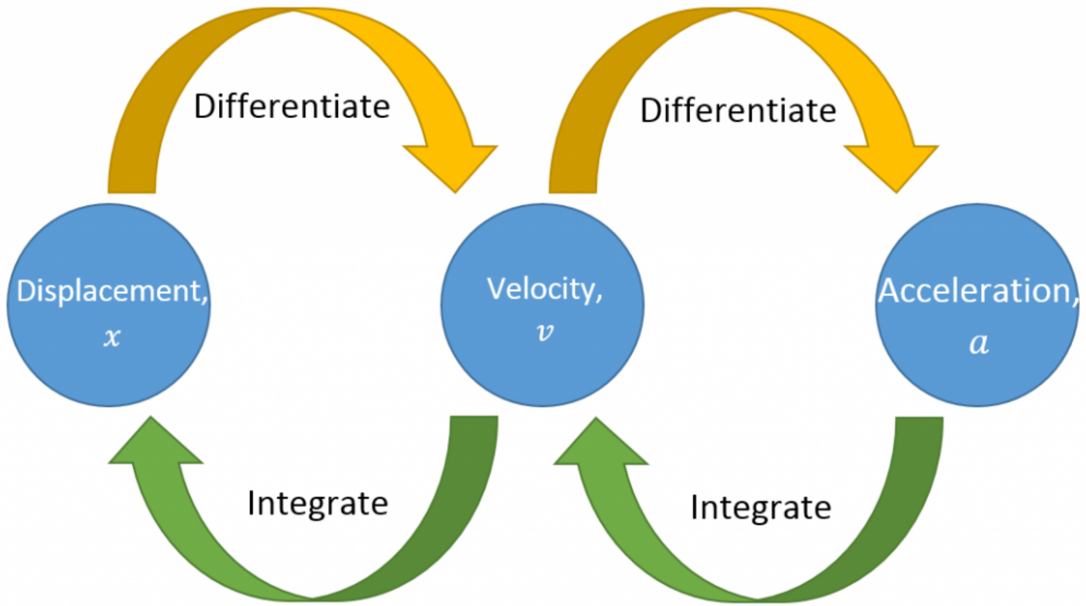
We next highlight the solutions that are used for pose detection in multiple practical systems!

Tracking Systems: In Depth Practical System Study

- Rotational tracking (3DoF) is always done with gyroscopes
  - Measure angular velocity
  - Part of IMU
- Different companies use different technologies to enable positional tracking (6DoF)
- Different tracking systems balance among different metrics
  - Cost
  - Ease of setup
  - Tracking volume
  - Controller tracking range

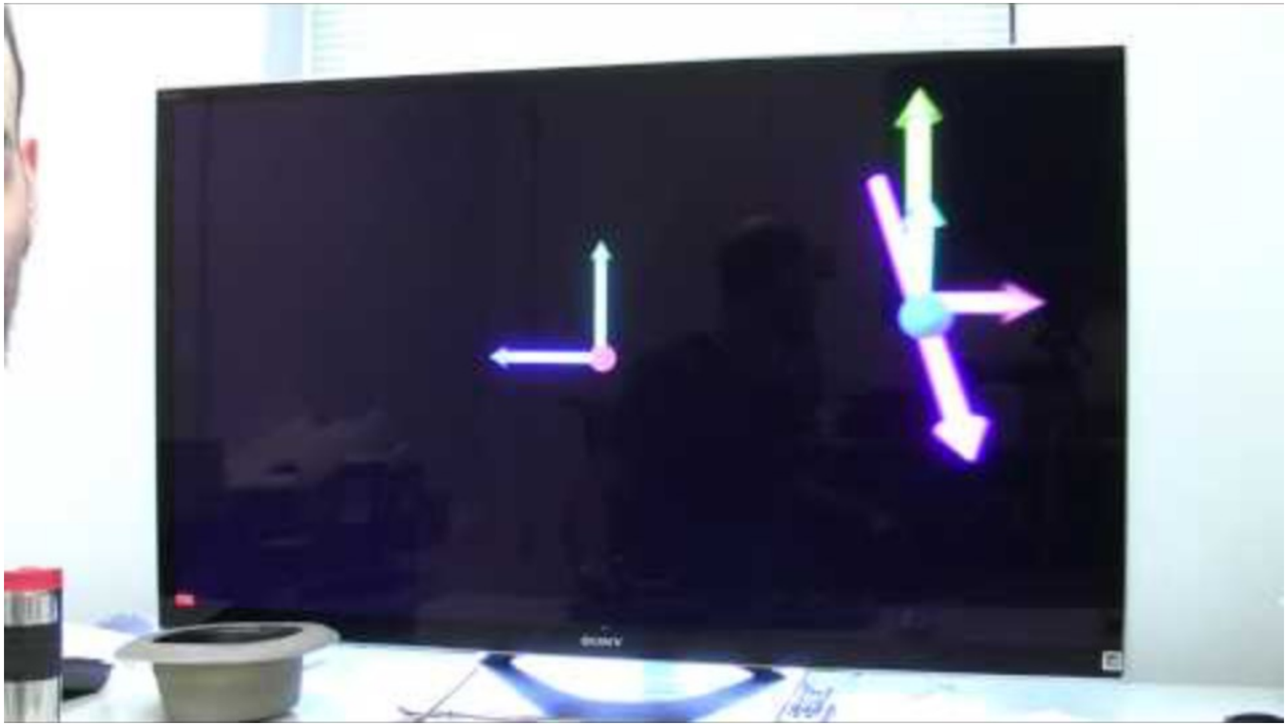
Common Base: Dead Reckoning

- In navigation, **dead reckoning** is the process of calculating current position of some moving object by using a previously determined position, or by using estimations of speed, heading direction, etc.
- The base solution is to use accelerometers in IMUs
  - Typically run at 1000 Hz
  - Take twice integration to derive position



From moment to moment it's how every VR headset and controller tracks itself.

So Why Anything Else is Needed?



So How Do VR Tracking Systems Correct?

The purpose of VR tracking systems is to correct this drift by providing a reference. Each tracking system does it differently, but the purpose remains the same.

We next go through several commercial tracking systems!

Constellation (Original Oculus Rift)

- Each tracked device has a pre-defined “constellation” of infrared LEDs hidden under the external plastic, which you can see highlighted in the image above. IR light is invisible to the human eye.
- The LED infrared light passes flawlessly through the plastic cover

Constellation (Original Oculus Rift)

- Sensors, which are basically cameras with filters to see only IR light, send frames to the user's PC over a USB cable at 60 Hz. The PC processes each frame, identifying the position of each IR LED, and thus the relative position of each object.

Pinning down a moving rigid body from n observed features is called "Perspective-n-Point (PnP) problem"

Constellation (Original Oculus Rift)

- The software can easily recognize which LEDs it's seeing because:
  - It knows the shape of the "constellation",
  - It remembers where the object was in the previous frame, and
  - It knows its direction of acceleration (from the accelerometer), and its rotation (from the gyroscope).
  - Each IR LED also blinks at a specific frequency to identify itself.

- To support fast motion, the Rift headset and Touch controllers wirelessly communicate with a wireless chip in the sensor every time they're about to pulse their LEDs. This allows the camera shutter to fire exactly as the LEDs do, and allows the exposure to be short.

- Constellation tracking is no longer used for headset tracking for the Rift S, but it is used on the headset onboard cameras to track the controllers.

Constellation Advantages/Disadvantages

- Advantages:
  - Low cost to integrate
  - High quality tracking
  - Works in most environments
- Disadvantages:
  - Each sensor has wired connection to PC
  - Large USB bandwidth causes issues with many motherboards
  - Sensors have limited vertical field of view

PlayStation VR

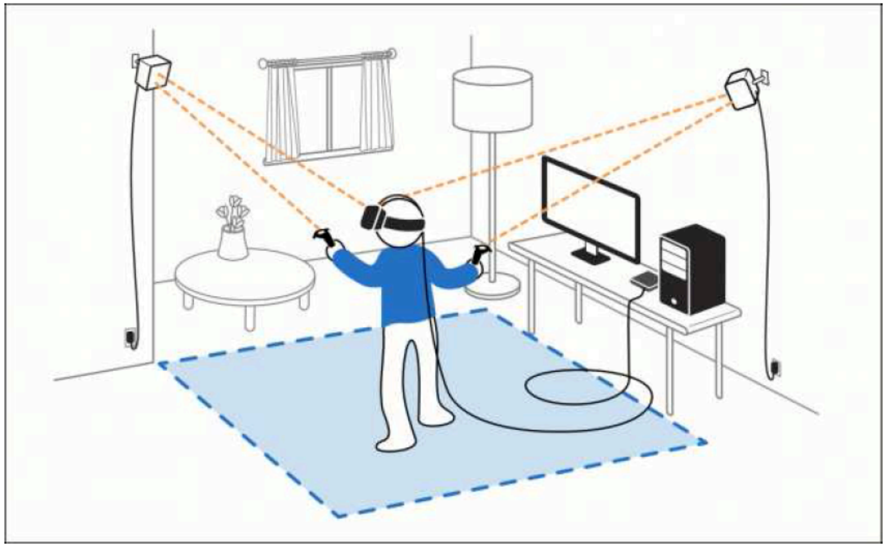
- PlayStation VR uses cameras too, but unlike the Rift the PSVR's tracking operates in the visual light spectrum.
- The PlayStation 4 camera bar contains two spaced out cameras. The camera unit is connected to the PlayStation, which uses the image data to track the blue strips of light on the headset and the orbs of light on the controllers.

PlayStation VR Advantages/Disadvantages

- PSVR Tracking Advantages:
  - Low cost to integrate
  - Leverages existing PS Move controllers
- PSVR Tracking Disadvantages:
  - Low quality tracking
  - No room scale support

SteamVR Lighthouse (HTC Vive)

- Unlike all the other systems it does not use cameras at all
- Lighthouse was designed from the start to enable room scale positional tracking without having to wire sensors back to the user's PC.
- Base stations ("Lighthouses") are placed at opposite top corners of the room. They do not communicate with the PC and they are not sensors. They emit wide angle two dimensional IR laser beams across the entire room. This is done 1 axis at a time, so left-right then top-bottom, repeatedly. Before each sweep they emit a powerful IR flash of light.



Lighthouse

SteamVR Lighthouse (HTC Vive)

- Note that infrared is not visible to human eye
- Each tracked device contains an array of IR photodiodes connected to a chip. This chip measures the time between the IR flash and being hit by the laser sweep for each axis. From this it can determine its position in the room.

Lighthouse Advantages/Disadvantages

- The unique value of Lighthouse is that it is a relatively simple design in theory. No complex computer vision algorithms or camera readings are involved, just the timing between laser sweeps.

- Lighthouse Advantages:
  - No PC connection to base stations required (just power),
  - High quality tracking,
  - Wide tracking volume.

- Lighthouse Disadvantages:
  - Relatively expensive to produce / integrate,
  - Usually requires mounting base stations on walls (otherwise the motors jitter),
  - Reflective surfaces in room cause glitches.

SLAM / Inside Out Tracking

- After the release of the Oculus Rift and HTC Vive, a lot of VR companies began to realize that the requirement to mount sensors or base stations in the room was putting many potential buyers off.

- Many newer and upcoming headsets use cameras built into the headsets themselves which perform “inside-out” tracking using computer vision algorithms. The specific type of algorithm used is called Simultaneous Location And Mapping (SLAM).

- SLAM algorithms work by noticing unique static features in the room. By comparing the rotation and acceleration from the accelerometer & gyroscope to how these features appear to move, the position of the headset can be determined.

SLAM / Inside Out Operation

Companies with SLAM Products

- Microsoft: Every Windows MR headset
- Google: Lenovo Mirage Solo
- HTC: Vive Focus
- Facebook: Oculus Quest and Oculus Rift S

- Google calls their tracking algorithm ‘WorldSense’ and Facebook calls theirs ‘Oculus Insight’.

- To track controllers, these systems operate in a similar way to Constellation — either with visible or infrared light. The cameras on the headset track LEDs under the plastic on the controllers.

SLAM Advantages/Disadvantages

- Advantages:
  - No external hardware required,
  - Very low cost,
  - Trivial setup,
  - If you could see your hands in real life, the headset can track the controllers.
- Disadvantages:
  - Doesn’t work in the dark,
  - Can’t track controller movement behind head or back accurately (because it uses normal light).
  - Can’t track controller movement when arm is between headset and controller.