Computer Vision and Mixed Reality

Tracking

Chapter 3

Pedro Mendes Jorge

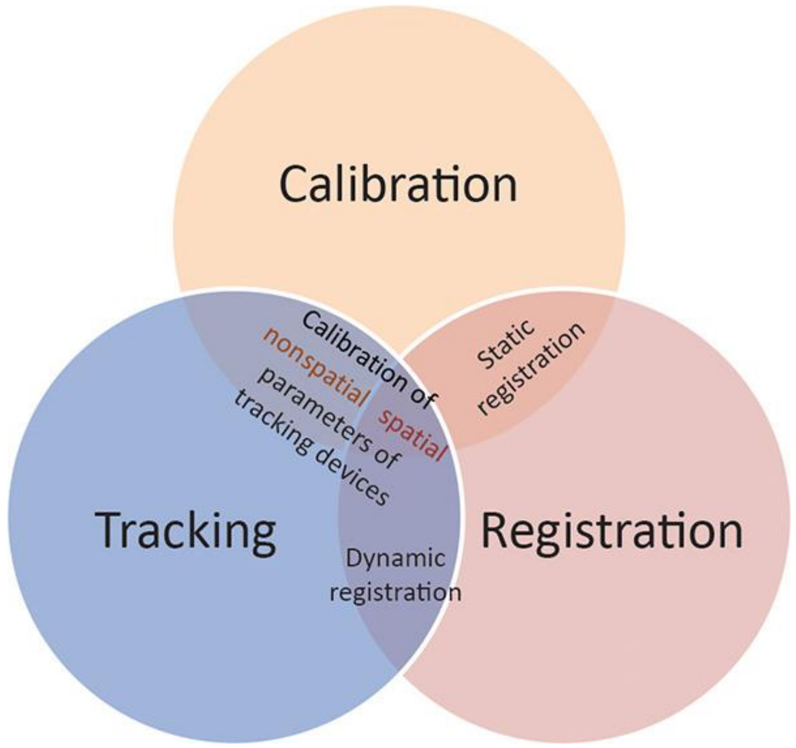
Tracking, Calibration and Registration

In AR, three important terms are associated with the measurement and alignment of objects — **tracking**, **calibration**, and **registration**

Tracking is responsible for dynamic registration

Calibration is the offline adjustment of measurements

Registration refers to the alignment of spatial properties



Tracking, Calibration and Registration

Tracking

- To display virtual objects registered to real objects in three-dimensional space, we must know at least the relative **pose — the position and orientation** of the AR display relative to the real objects

- Pose measurements must be continuously updated (real time tracking)

- Complete “3D tracking” is related with the estimation of the six-dimensional pose (position and orientation) of real entities

Tracking, Calibration and Registration

Calibration

- Calibration is the process of comparing measurements made with two different devices, a reference device and a device to be calibrated
- The objective is to determine parameters for using the device to be calibrated to deliver measurements on the known scale
- For AR, we need to calibrate the components of the AR system, especially the devices used for tracking
- Depending on the measurement system, calibration may be done just once in a device’s lifetime, every time before commencing operation, or concurrently with tracking

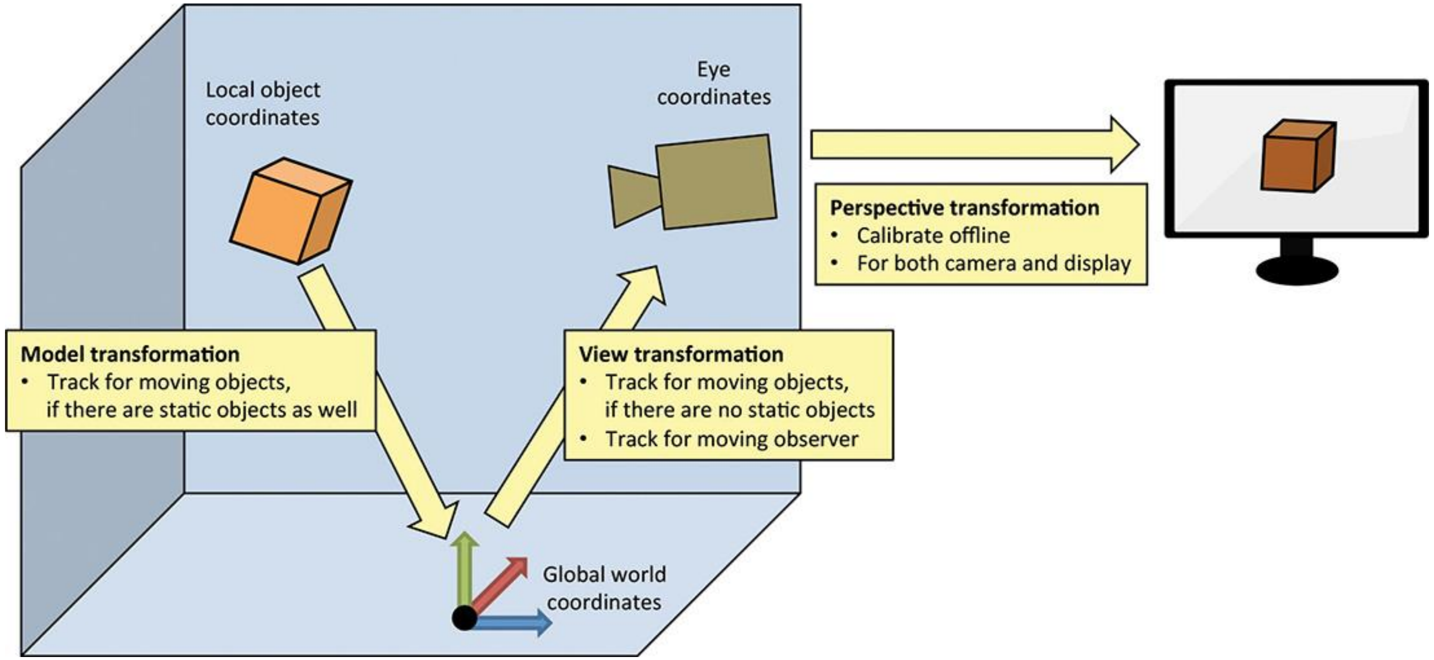
Tracking, Calibration and Registration

Registration

- Registration in AR refers to the alignment of coordinate systems between virtual and real objects
- This requires tracking of the user's head or of the camera providing a video background (or both)
- **Static registration**, when the user or the camera is *not* moving, requires calibration of the tracking system to establish a common coordinate system between virtual and real object.
- **Dynamic registration**, when the user or the camera is moving, requires tracking.

Coordinate Systems

- Model Transformation
- View Transformation
- Projective Transformation



Coordinate Systems

- The **model transformation** describes the pose of moving objects in a static environment
- The **view transformation** describes the pose of a camera, a tracking sensor, or a display in an environment
- The **perspective transformation** describes the mapping from eye coordinates to screen coordinates
- Both the model and view transformations can be tracked, enabling registration.

Coordinate Systems

- Registration implies that the cumulative effect of these transformations must be matched between the real and the virtual

- How we deal with the individual transformations depends on the configuration of the AR system and the tracking technology

- Certain parameters can be calibrated offline, whereas other parameters change on a frame-by-frame basis and need tracking

Coordinate Systems

Model Transformation

- The model transformation describes the relationship of 3D local object coordinates and 3D global world coordinates
- The model transformation determines where objects are placed in the real world
- For every moving real object in the scene with which we want to register virtual information, we must track its model transformation

Coordinate Systems

View Transformation

- The view transformation describes the relationship of 3D global world coordinates and 3D camera coordinates
- Most AR scenarios allow an observer to move in the real world and tracking the view transformation is the most important objective
- AR typically requires a separate viewing transformation for the camera and the display of the user

Coordinate Systems

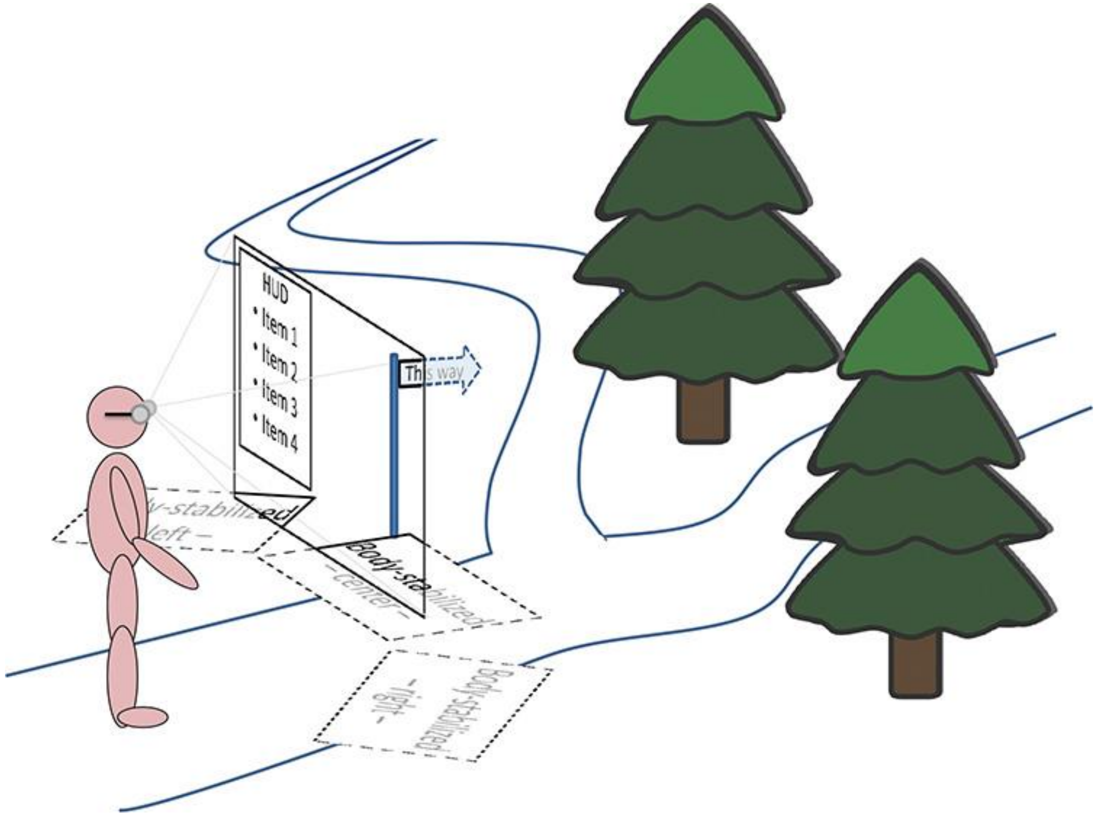
Projective Transformation

- The projective transformation describes the relationship of 3D camera coordinates and 2D device coordinates
- The content of the view *frustum* (truncated pyramid) is mapped to a unit cube and then projected onto the screen by dropping the Z component and applying a viewport transformation
- The projective transformation is usually calibrated offline

Coordinate Systems

Frames of Reference

- Virtual information can be fixed with respect to the global world, a (potentially moving) object, or a person’s view (the AR screen)
- *World-stabilized*, *object-stabilized* (or in the special case, *body-stabilized*), and *screen-stabilized* information



Characteristics of Tracking Technology

- Physical Phenomena
  - Measurements can exploit **electromagnetic radiation** (including visible light, infrared light, laser light, radio signals, and magnetic flux), **sound**, **physical** linkage, **gravity**, and **inertia.**
  - Specialized sensors are available for each of these physical phenomena

- Measurement Principle

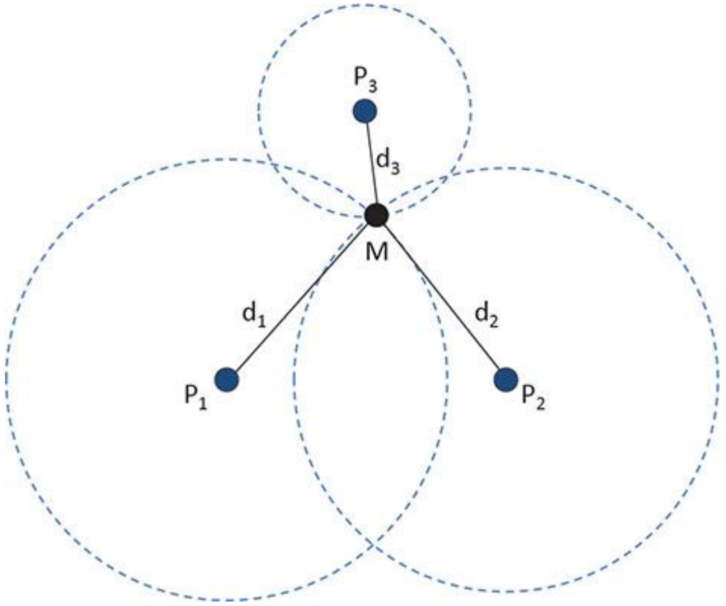
  It can be measured **signal strength**, **signal direction**, and **time of flight** (both absolute time and phase of a periodic signal) – TOF.

  TOF measurements require some form of secondary communication channel to confirm clock synchronization between sender and receiver.

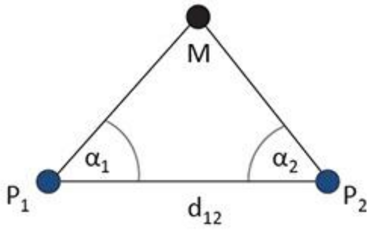
Characteristics of Tracking Technology

Measured Geometric Property

- It can be measured either distances or angles



Trilateration



Triangulation

Characteristics of Tracking Technology

Sensor Arrangement

- A common approach is to use **multiple sensors** together in a known rigid geometric configuration, such as a stereo camera rig
- Sometimes, it is important to arrange **three sensors orthogonally** to measure vector-valued quantities, such as acceleration in three fundamental directions
- Multiple-sensor configurations need **sensor synchronization** to ensure simultaneous acquisition of measurements
- The process of combining multiple sensor inputs to obtain a more complete or more accurate measurement is referred to as **sensor fusion**

Characteristics of Tracking Technology

Signal Sources

- Sources can be either **passive** or **active**
- **Passive sources** rely on natural signals present in the environment, such as natural light or the Earth's magnetic field
- When no external source is apparent, such as in inertial sensing, the signaling method is described as **sourceless sensing**
- **Active sources** rely on some form of electronics to produce a physical signal

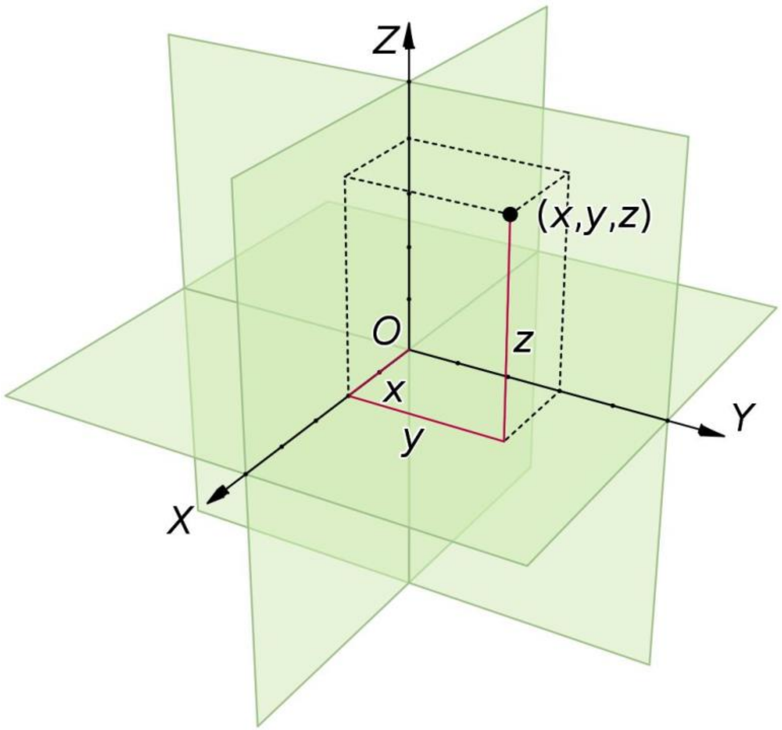
Characteristics of Tracking Technology

Measurement Coordinates

- Tracking measures physical quantities relative to a given coordinate system
- Humans are spatial beings, we see, hear, interact with, and move through our physical surroundings in **three dimensions**
- Our perceptual mechanisms are clearly optimized to process information in this form

Defining Position and Orientation in 3D

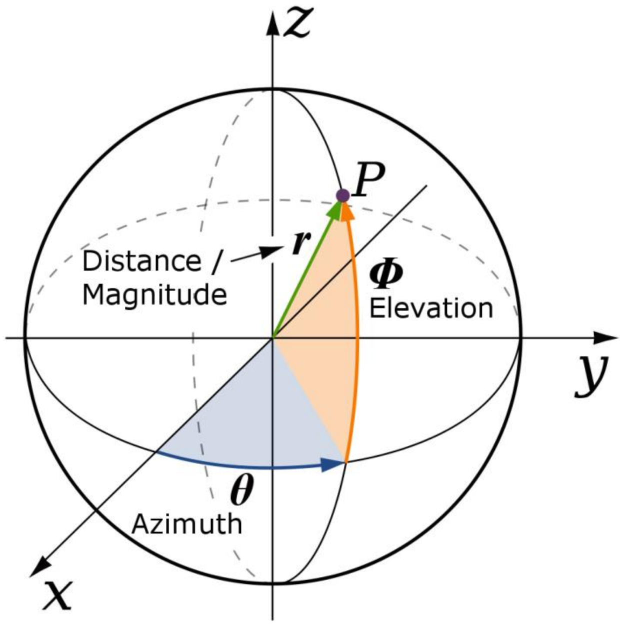
Position



Cartesian Coordinates - The standard coordinate system is used to define a virtual space. Within this system, each point is uniquely specified using three numerical coordinates (x,y,z) representing specific distances (measured in the same unit of length) from three mutually perpendicular planes.

Defining Position and Orientation in 3D

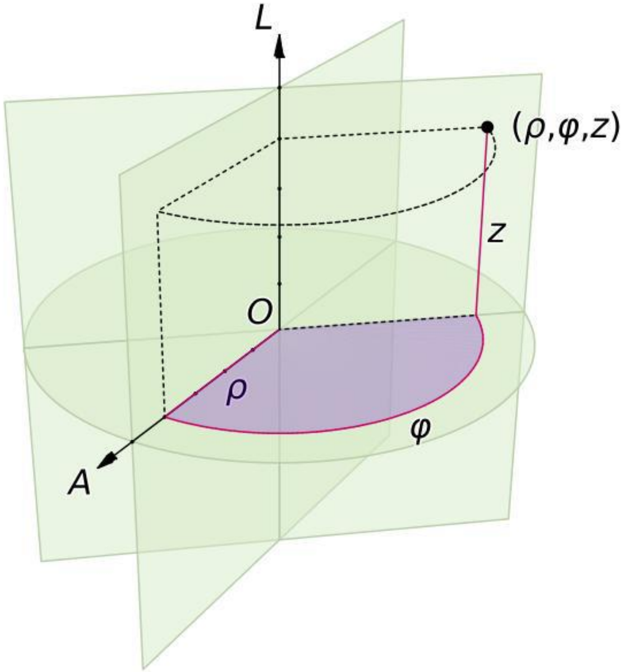
Position – Angular Method



Spherical Polar Coordinates - Often used to define the location of objects and features relative to the user's position. The system is based on fundamental planes bisecting a sphere and is composed of three elements: azimuth, elevation, and the distance (magnitude or range).

Defining Position and Orientation in 3D

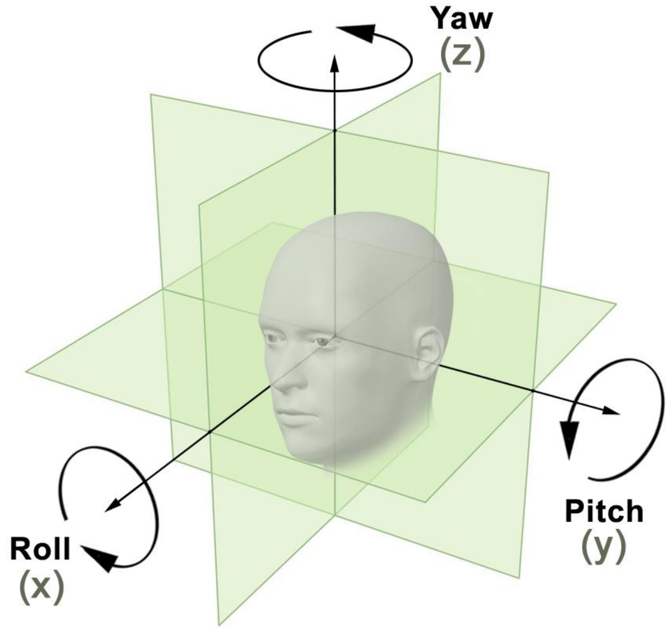
Position – Angular Method



Cylindrical Coordinates - Used extensively in the creation of still-image mosaics for "immersive" panoramas or 360-degree video backgrounds. Allows for the precise mapping and alignment of multiple images for overlap and edge stitching, with all points being the same distance from a central reference axis.

Defining Position and Orientation in 3D

Defining Orientation and Rotation



It is common to see position (x,y,z) and orientation (roll, pitch, and yaw) generically referred to as six degrees of freedom and frequently abbreviated as 6 DOF.

Characteristics of Tracking Technology

Degrees of Freedom

- In measuring systems, a **degree of freedom** (DOF) is an independent dimension of measurement
- Registering real and virtual objects in three-dimensional space usually requires determining the pose of objects with six degrees of freedom (6DOF): three degrees of freedom for position and three degrees of freedom for orientation
- Certain sensors and technologies deliver only 3DOF orientation (for example, a gyroscope), or only 3DOF position (for example, a single tracked LED), or just one or two specific degrees of freedom (for example, an odometer)

Characteristics of Tracking Technology

Measurement Coordinates

Global versus Local

- **Local** implies a smaller-scale coordinate system established by the user, possibly employing an *ad hoc* approach. For example, we may measure position relative to a corner of the room we are in.
- **Global** implies worldwide measurements (or at least very wide area, such as entire-city scale), which is still relative, but with respect to the entire planet. For example, a compass measures heading relative to the Earth’s magnetic field.

Characteristics of Tracking Technology

Measurement Coordinates

Absolute versus Relative

- **Absolute** measurement (e.g., measurement of a moving object's pose) implies that the reference coordinate system is set in advance
- **Relative** measurement means that the reference coordinate system is established dynamically (e.g., relative to a previous pose)

- Relative measurements are more difficult to exploit in AR, because registration of real and virtual usually is expected to be fixed and not continuously changing

Characteristics of Tracking Technology

Spatial Sensor Arrangement

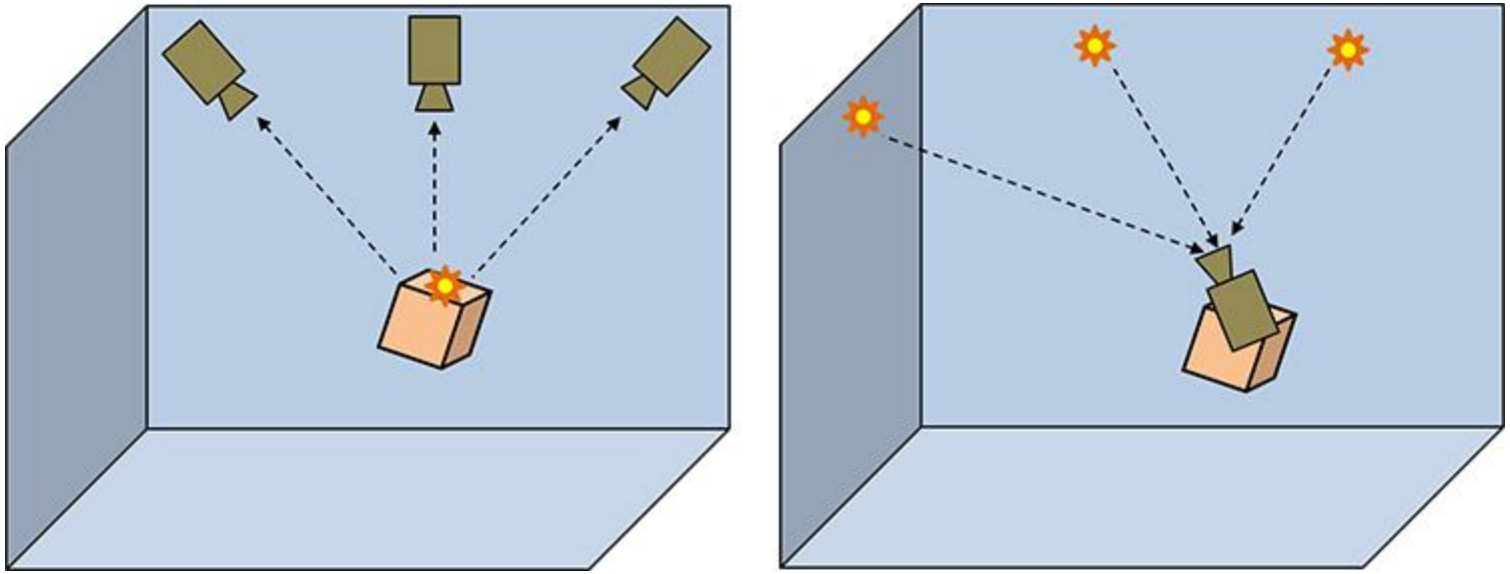
Two fundamental types of spatial arrangements for tracking systems exist: **outside-in** versus **inside-out**

Outside-in tracking uses sensors mounted stationary in the environment

Inside-out tracking uses sensors mounted to a mobile or body-worn device

Characteristics of Tracking Technology

Spatial Sensor Arrangement



Outside-in　　　　Inside-out

Characteristics of Tracking Technology

Workspace Coverage

- Ideally, users would like to freely roam through an arbitrary large environment, both indoors and outdoors, and be free of encumbering artifacts, such as tethers for electronic devices
- While a larger range is clearly desirable for mobile applications, a tradeoff between range and accuracy is usually required
- Tracking systems, however, typically rely on some sort of infrastructure composed of active devices, such as outside-in tracking systems, or passive targets, such as fiducial targets mounted in the environment or carried by the user

Characteristics of Tracking Technology

Measurement Error

- The **accuracy** of a measurement is determined by how close the measurement is to the true value of the quantity being measured

- The **precision** of a measurement is defined as how closely a number of measurements of the same quantity agree with each other

- The **resolution** of a sensor is the minimum difference that can be discriminated between two measurements

Characteristics of Tracking Technology

Temporal Characteristics

There are two important temporal characteristics of tracking systems: **update rate** and **latency**

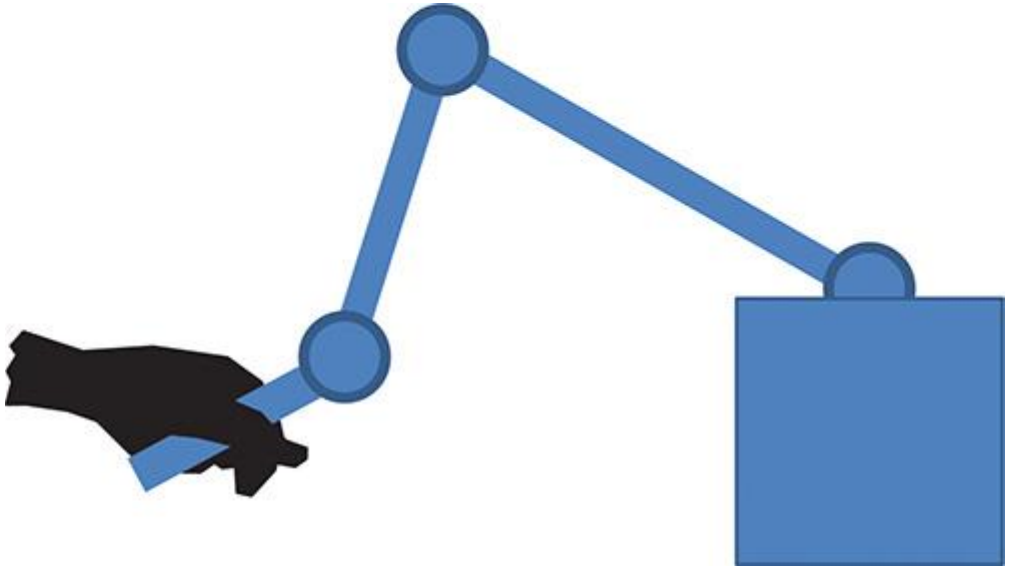
Update rate (or temporal resolution) is the number of measurements performed per given time interval

Latency is the time it takes from the occurrence of a physical event, such as a motion, to a corresponding data record becoming available to the AR application

Latency is arguably the more critical one for real-time applications such as AR, because it determines more directly how much dynamic error is introduced on a system level

Stationary Tracking Systems

Mechanical Tracking



Example of a mechanical articulated arm using an arrangement of limbs and joints (here: three each). The angles at the joints are sensed. Such a setup can also provide force feedback.

Stationary Tracking Systems

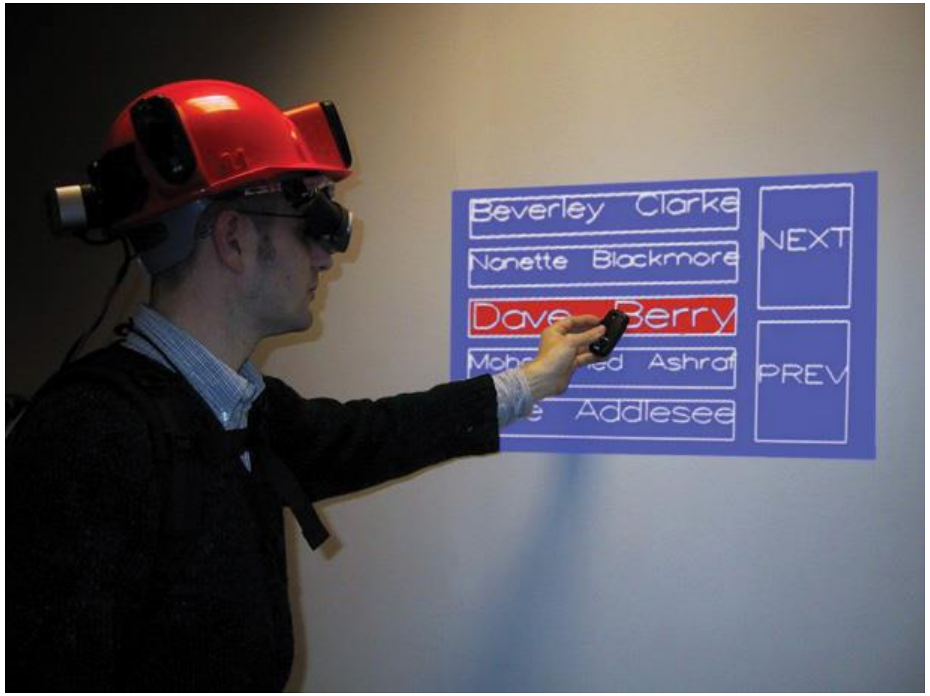
Electromagnetic Tracking

The Razer Hydra is a short-range magnetic tracking device designed for desktop use. The 6DOF pose of two handheld joystick controllers is tracked relative to the spherical base.

Stationary Tracking Systems

Ultrasonic Tracking

- Ultrasonic tracking measures the time of flight of a sound pulse traveling from source to sensor.



The AT&T Bat system was an ultrasonic outside-in tracking system. Three Bat emitters are mounted on a helmet. One emitter in the hand serves as a pointer. Receivers are mounted to the ceiling, and a time-sharing approach allows tracking over wide areas, such as an entire building. Courtesy of Joseph Newman.

Mobile Sensors

Tracking systems for AR tend to be mobile

However, AR users roaming an unconstrained environment—in particular, outdoors—cannot expect control over the physical infrastructure and cannot rely on constant quality of wireless services

Consequently, both sensing and computation for tracking must be performed locally on the mobile device, usually without the aid of infrastructure in the environment

This limits the applicable techniques to the ones that can operate with mobile sensors and limited processing power

Mobile Sensors

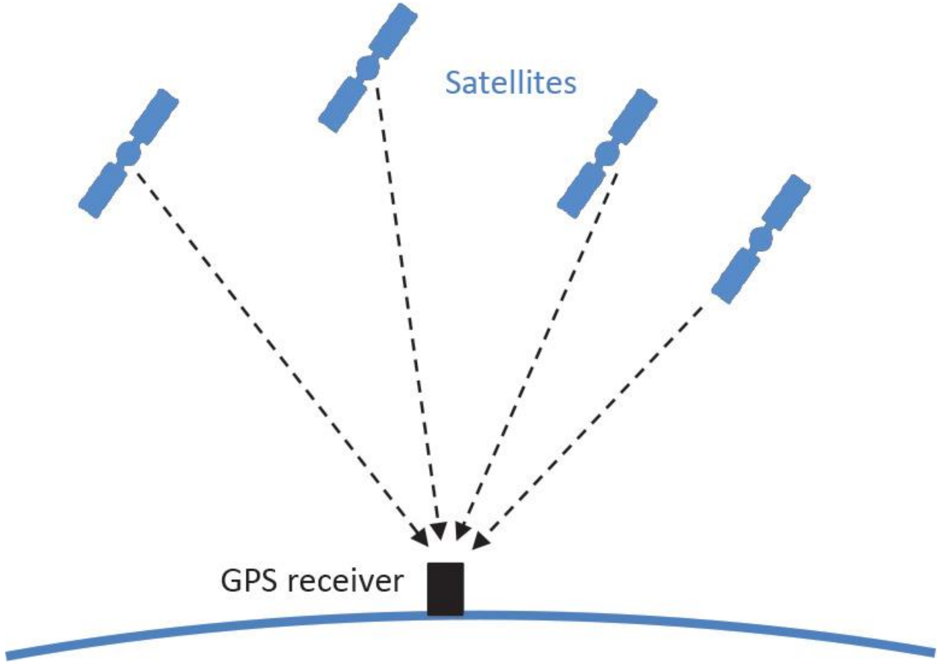
Modern mobile devices such as smartphones or tablets are equipped with an array of sensors

While the performance of these sensors is strongly limited by external constraints, the sensors present a significant opportunity, because they are integrated into inexpensive devices and continuously available

Mobile Sensors

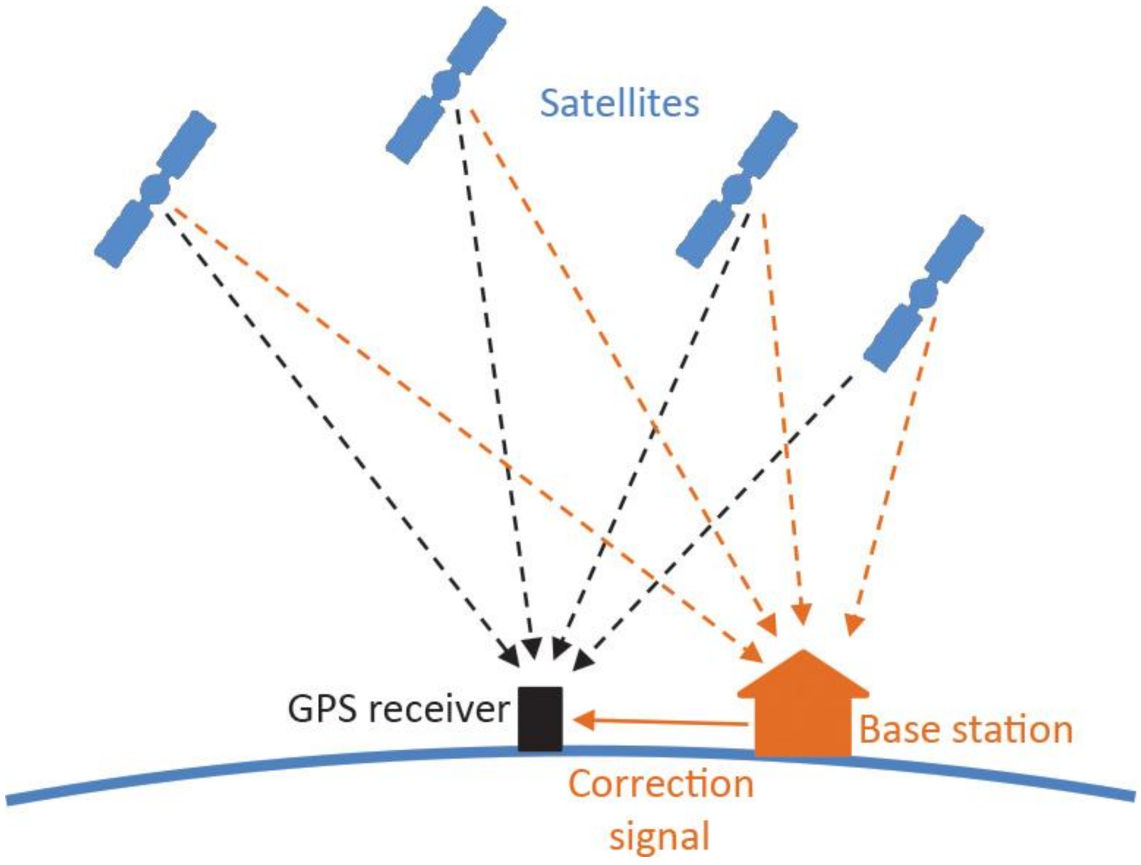
Global Positioning System

Global navigation satellite systems—in particular, the Global Positioning System (GPS) developed by the United States—measure the time of flight of coded radio signals emitted by satellites in Earth orbit, essentially representing a planet-sized inside-out system



Mobile Sensors

Differential GPS



Mobile Sensors

Wireless Networks

- Existing wireless network infrastructures, such as **Wi-Fi, Bluetooth, and mobile phone networks**, can be used to determine one's position
- Position can be obtained by measuring the **signal strength** of base stations to estimate distances and apply geometric reasoning based on **trilateration**
- Low-power, low-cost **beacons** based on Bluetooth are a possible dedicated infrastructure for **indoor positioning**—for example, in retail applications
- Given that **GPS, Wi-Fi, and cellular radio** capabilities are usually simultaneously available in **mobile devices**, such systems now commonly **combine information** from all three sources for improved coverage, speed, and accuracy of position measurement (**data fusion**)

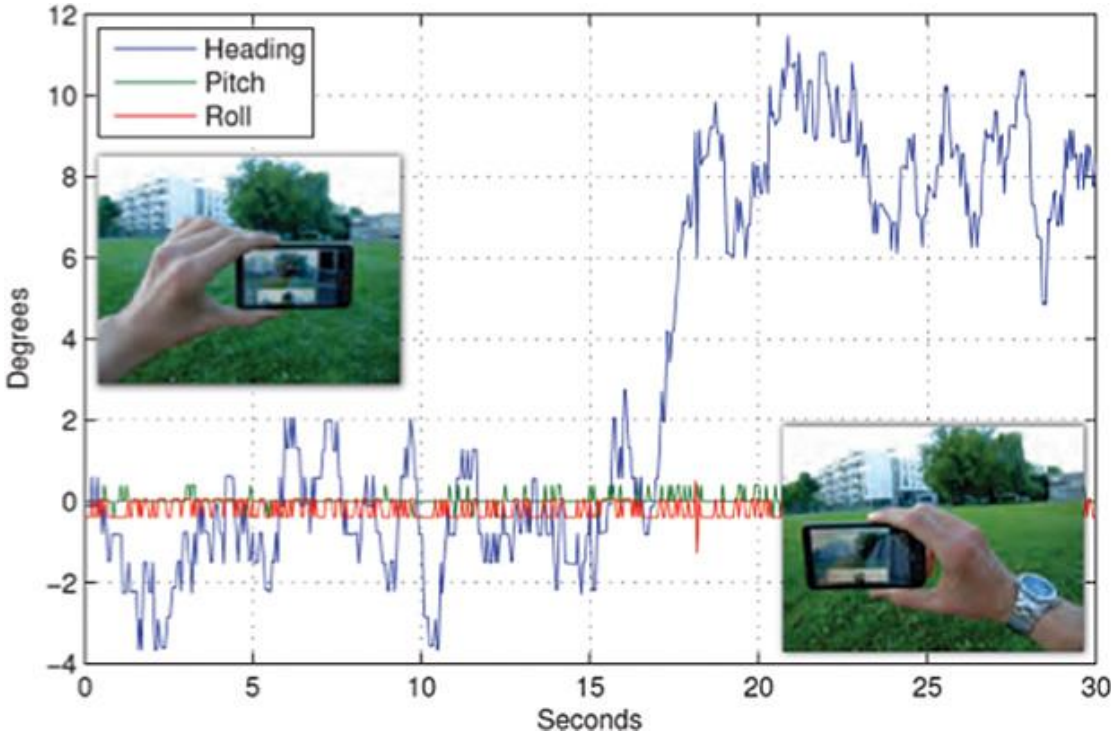
Mobile Sensors

Magnetometer

- A magnetometer, or electronic compass, measures the direction of the Earth's magnetic field to determine the bearing relative to the magnetic north
- In practice, magnetic measurements are often unreliable, since they are subject to distortion from local magnetic fields, such as those created by electric and electronic equipment.

Mobile Sensors

Magnetometer



Every modern smartphone contains a magnetometer, but precision of that sensor alone is often very poor. The image shows the heading error over time as a metallic watch worn on the user’s right hand comes close to the device. Courtesy of Gerhard Schall.

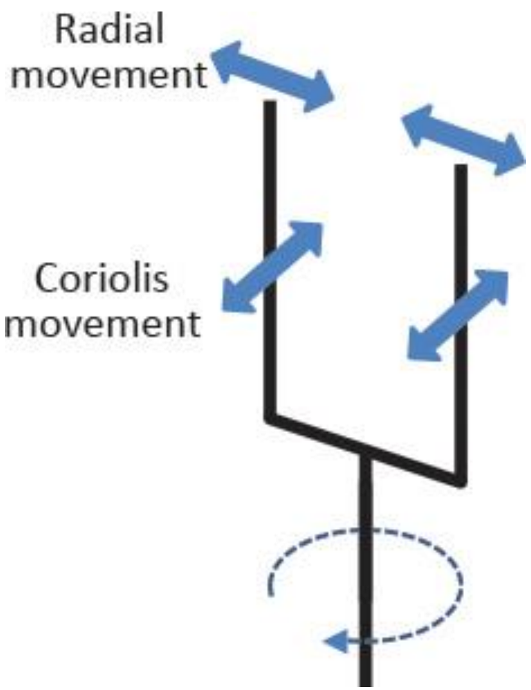
Mobile Sensors

Gyroscope

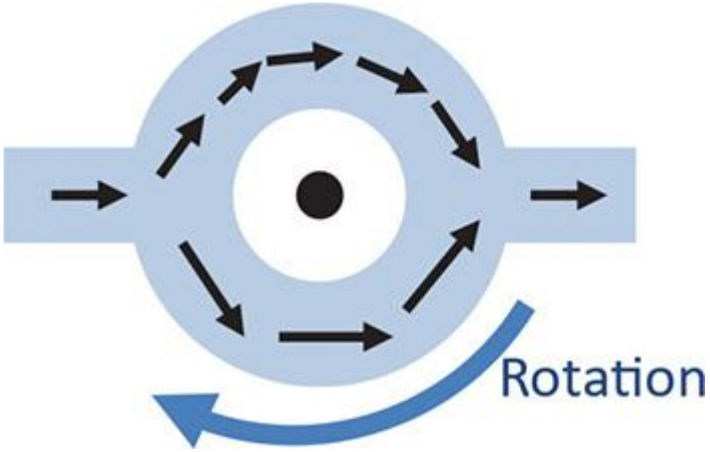
- An electronic gyroscope is a device for measuring rotational velocity.
- With numerical integration, the orientation can be computed.
- Three orthogonal gyroscopes are usually combined in a micro-electromechanical system (MEMS) to deliver a full 3DOF orientation measurement.
- Laser or fiber-optic gyroscopes, like those used in aviation, measure angular acceleration based on interference of light (*Sagnac* effect) observed at the end of a looped fiber-optic coil.

Mobile Sensors

Gyroscope



MEMS gyroscopes measure the out-of-plane motion of a mass vibrating in a plane orthogonal to the rotation axis.

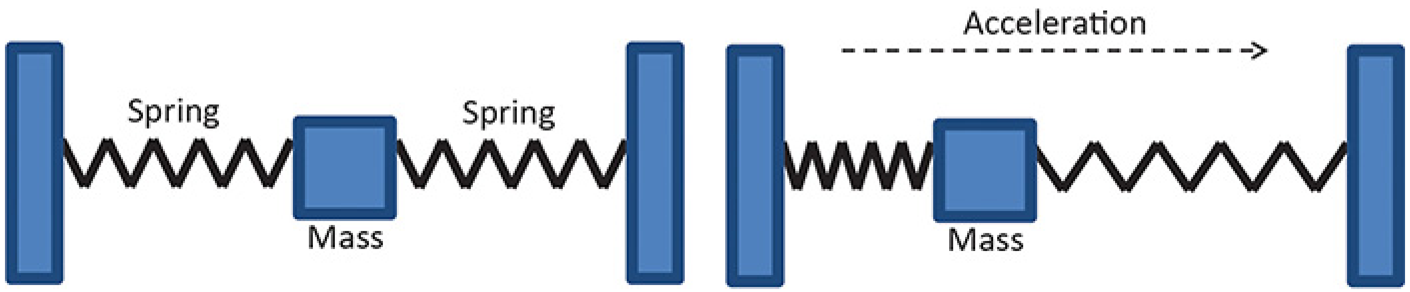


Laser gyroscopes based on the *Sagnac* effect in optical fiber are highly precise.

Mobile Sensors

Linear Accelerometer

- Acceleration exerts a force on small mass and the resulting displacement is measured along each major axis separately
- After subtracting the effects of gravity and integrating twice numerically, position can be computed from the acceleration measurements (determine position relative to a starting point)



A one-dimensional linear accelerometer measures the displacement of a small mass suspended between springs as the sensor accelerates.

Mobile Sensors

Odometer

- An odometer is a device frequently used in mobile robotics or vehicles to incrementally measure the distance traveled over ground.
- A mechanical or opto-electrical wheel encoder determines the number of turns taken by a wheel on the ground.
- Multiple encoders allow observing turns of the device.
- For example, inexpensive odometers were used in the traditional computer mouse to observe the rotation of the ball inside the mouse.

Optical Tracking

- Mobile sensors are important because of their **mobility**, but their **accuracy is usually not sufficient** to achieve the high-quality registration required in AR.

- Digital cameras are small, are cheap, and provide very rich sensory input—literally millions of independent pixels are acquired at once.

- **Optical tracking is easily one of the most important physical tracking principles used today for AR**.

Optical Tracking

Digital cameras

- Digital cameras are based on either complementary metal oxide semiconductor (**CMOS**) technology or charge-coupled device (**CCD**) chips.
- The lens of the camera plays an important role in determining the characteristics of the camera.

  For example, industrial cameras with large lenses provide much better quality than camera phones with tiny lenses of 1–2 mm diameter
- Thus, the **type of sensor**, the **lens**, and the type of **shutter** (e.g., global shutter, rolling shutter) determine the physical performance of the camera.

Optical Tracking

Model-Based versus Model-Free Tracking

Using images obtained from a camera requires comparing these images to some reference model.

- **Model-based tracking** – A reference model is obtained prior to starting the tracking system.
- **Model-free tracking** – A temporary model is acquired on the fly during the tracking.

Optical Tracking

Model-free tracking

- On-the-fly techniques such as **simultaneous localization and mapping** (SLAM) can combine 3D tracking with 3D scanning.
- Model-free tracking can determine the pose only relative to the starting point, **similar to an odometer**.
- Virtual objects in AR must be placed **spontaneously** and cannot be pre-registered to the real world.
- Recently, systems that **combine** the advantages **of model-based and model-free** tracking have become commercially available (Vuforia).

Optical Tracking

Illumination

- Passive Illumination
  - Light sources are not an integral part of the tracking system
  - Illumination comes both from natural light sources, in particular the sun, and artificial light sources, such as ceiling lights
- Active Illumination
  - Overcomes the dependence on external light sources in the environment by combining the optical sensor with an active source of illumination
  - Because active illumination in the visible spectrum changes how the user perceives the environment, infrared illumination can be used

Optical Tracking

Illumination – Structured Light

- Structured light **projects a known pattern** onto a scene (active illumination)
- The source of the structured light can be a **conventional projector** or a **laser light** source
- The observed reflections are picked up by a camera and used to detect the geometry of the scene and the contained object
- Laser ranging (LIDAR – *Light Detection And Ranging*) determines the time of flight taken by a laser pulse reflected from a surface

Optical Tracking

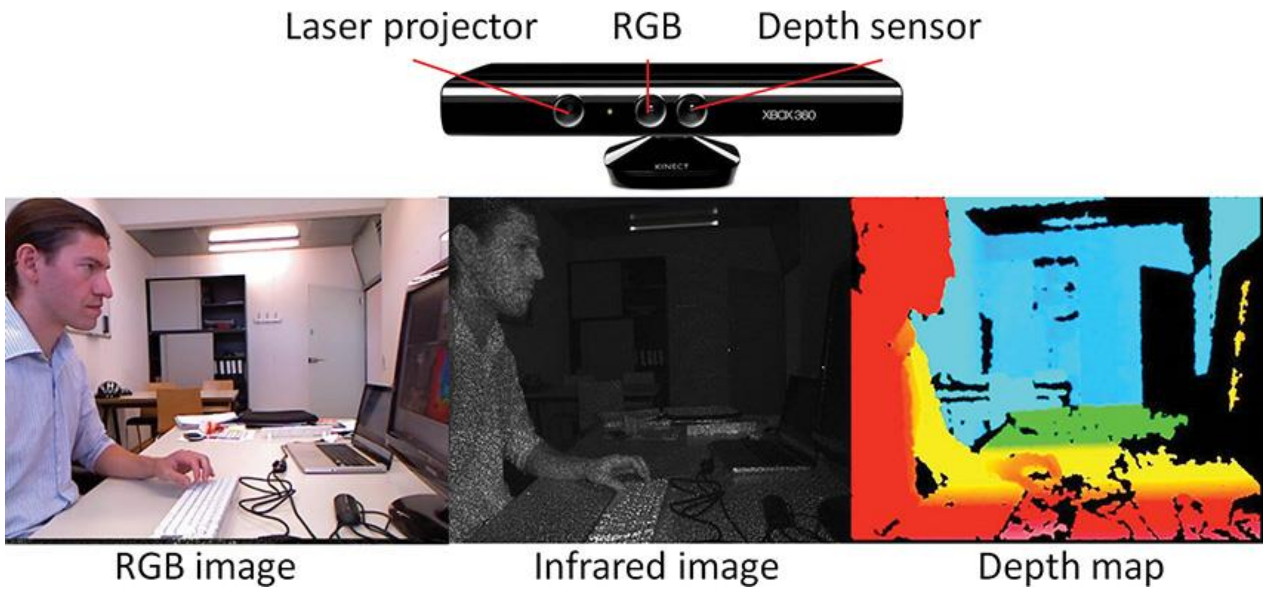
Illumination – Structured Light - Laser Ranging

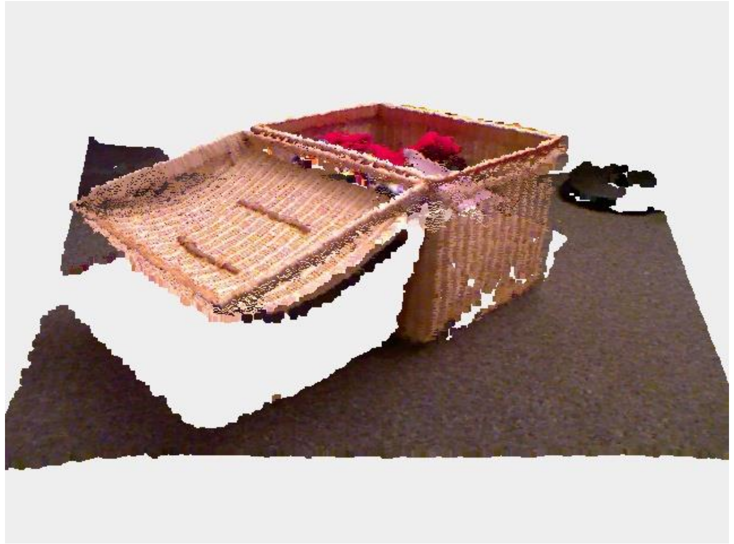
- In its simplest form, only a **single distance** is measured
- With the addition of a **rotating mirror**, the laser can be steered in either **one or two dimensions**—a configuration sometimes called a **laser scanner**.
- One-dimensional laser scanners are frequently mounted on mobile robots as input for autonomous navigation
- Stationary two-dimensional laser scanners delivering range images are used for 3D object reconstruction

Optical Tracking

Illumination – Structured Light - Laser Ranging

RGB-D camera (***D**epth*) devices are very appealing for AR, delivering images and geometric information about the scene, already registered.





The first-generation Microsoft Kinect used a structured light pattern projected with an infrared laser, while the second-generation uses a time-of-flight camera.

Optical Tracking

Markers versus Natural Features

Tracking targets can be classified into those with **natural features** and those with **artificial features**.

Tracking the natural environment is described as **natural feature tracking**.

Artificial pattern if often called a **marker** or **fiducial.**

Physical objects or digital models are known and need to be ease to detect and track.

Optical Tracking - Markers

Markers are needed when is **not reliable** to identify features of the target objects due to:

- objects may be **uniformly colored with little or no texture**, such as a white wall;
- objects may have a **high specularity**, so that their appearance is extremely unstable when moved relative to a camera;
- objects may have **repetitive textures**, such as a facade with identical windows.

Optical Tracking - Markers

Markers are **known patterns** placed on the surfaces of target objects or known trackable shapes attached to the target objects.

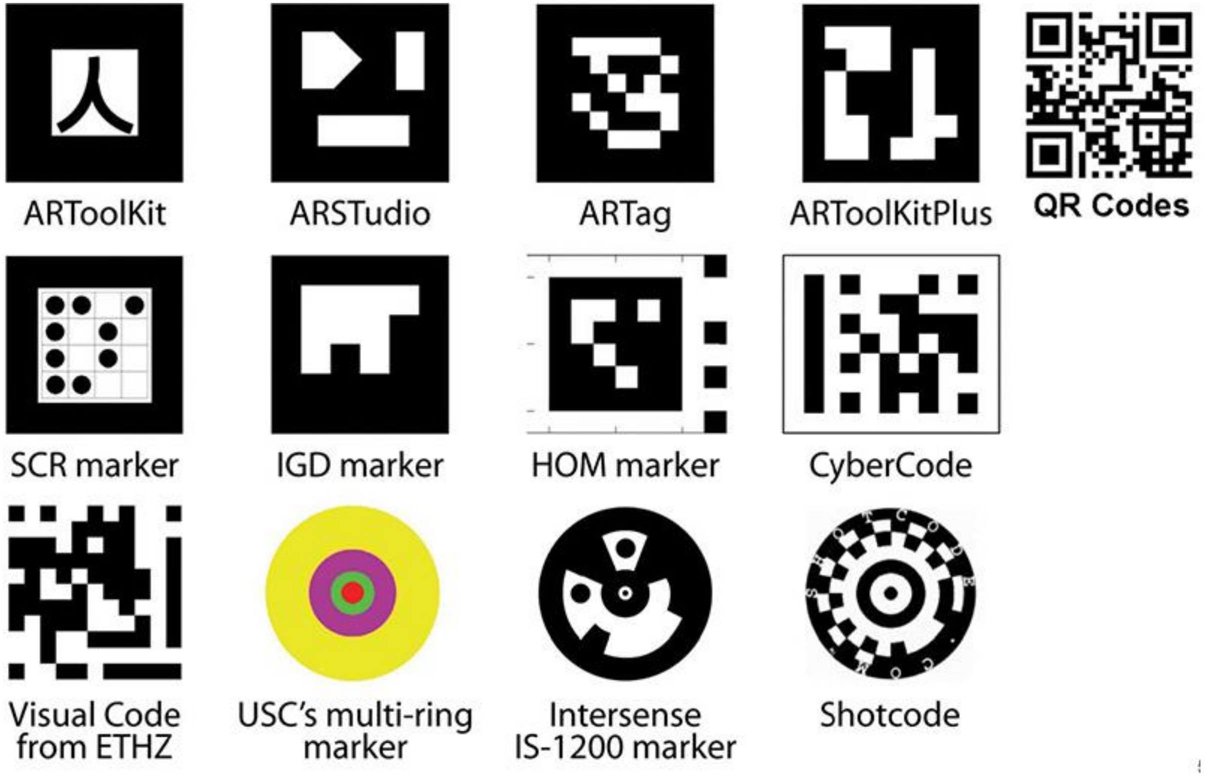
Markers are designed to make **detecting** their appearance in the image as **easy and reliable as possible**.

Markers are often designed as **black-and-white shapes** providing **good contrast** and is independent of how the camera handles color internally.

Markers can be **easily manufactured with an office printer**.

Optical Tracking - Markers

Examples



XI WANG [2019] - Augmented Reality (AR) Equipped Composites Repair

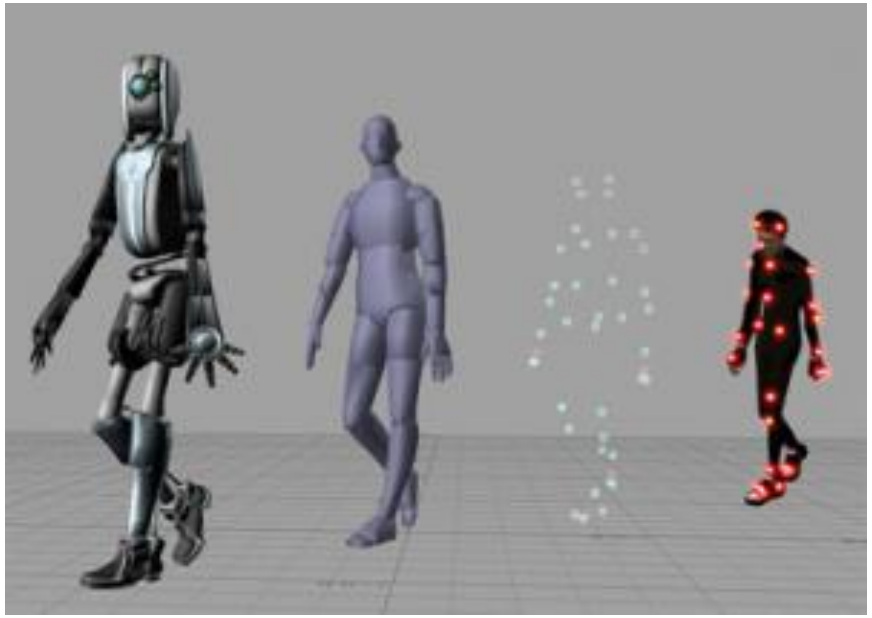
Optical Tracking - Markers

Other marker designs involve **spheres** attached to target objects in a **rigid configuration**.

Spheres have the advantage that they always **project to a disc shape in the image**, independent of the vantage point.

Because a **single spherical** marker identifies only a **single point**, at least three such markers are required.

Retro-reflective foil can be used to manufacture markers.



Optical Tracking - Natural Features

Natural feature tracking typically requires better image quality and more computational resources.

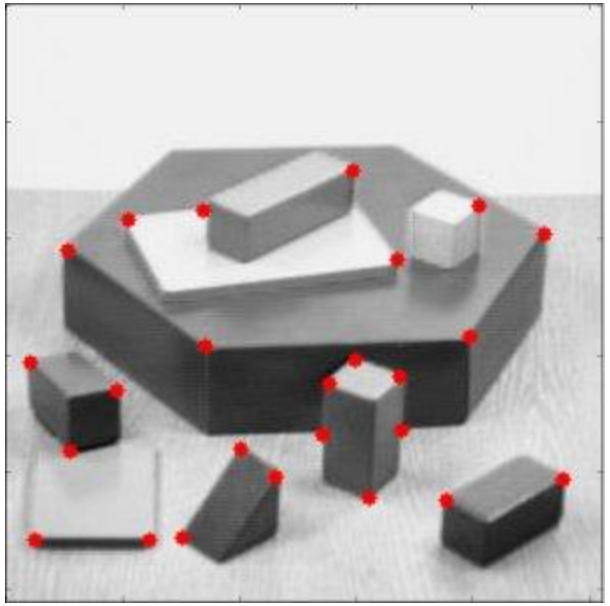
The most frequently used natural features are so-called **interest points** or **keypoints**, which are salient point features on a target object.

Interest points must be **easily found**, and their location on the object should **remain stable** under a changing vantage point.

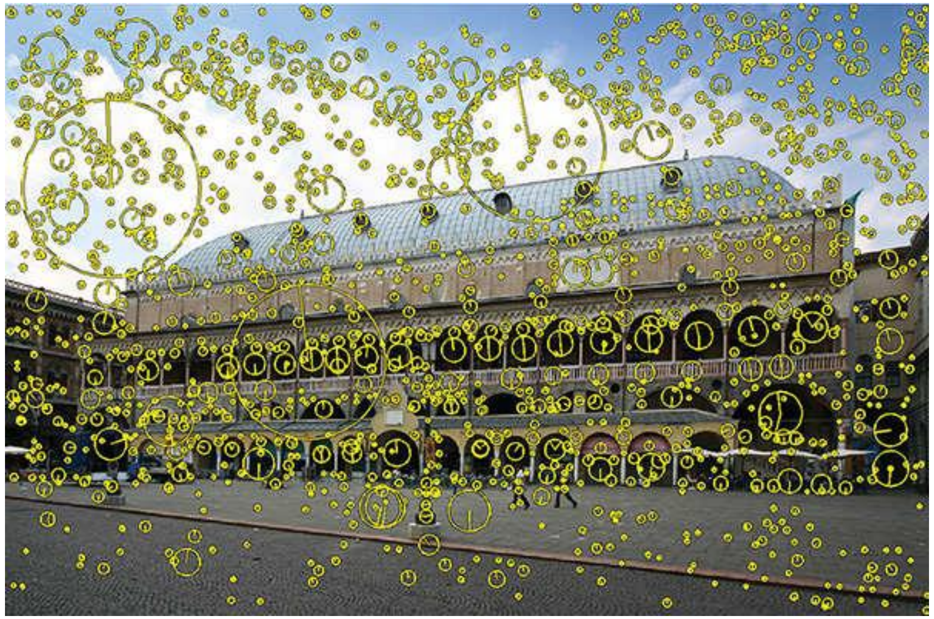
In practice, the use of interest points requires a **sufficiently dense and irregular surface texture**.

Optical Tracking - Natural Features

Interest point examples



Shi-Tomasi Corner Detector
[OpenCV docs]



SIFT interest points detected in an outdoor scene.

Optical Tracking - Natural Features

Objects that do not possess much texture can be tracked using **edge features**, assuming that their outline is easily observable.

A single edge hardly allows a unique identification without additional knowledge, and **multiple edges** must be jointly interpreted for reliable target detection.

It is also possible to compare the camera image using whole-image alignment to **keyframes** taken at specific vantage points.

Unfortunately, this approach is difficult to scale to larger environments.

Optical Tracking

Target Identification

To track multiple objects or track a mobile user in a wide area, identification of targets becomes a major issue of optical tracking.

However, a tradeoff must be made between the number of objects that can be identified and the reliability of the identification.

Supporting a larger set of objects necessarily means their appearance will become more similar and can be more easily confused.

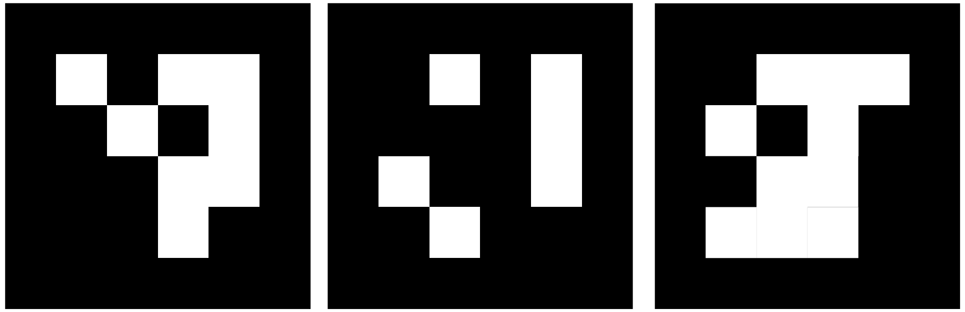
Optical Tracking - Target Identification

Marker Target Identification

Barcodes embedded in marker designs have a clear information payload, expressed in the number of bits encoded in the barcode.

Two bits are required to determine a unique orientation of the pattern.

Typical configurations use between 6 and 12 bits for the ID, allowing for a few thousand unique markers.



ArUco markers 4x4: ID 0; 50; 100

Optical Tracking - Target Identification

Marker Target Identification

The number of targets composed of **spherical markers** that can be reliably distinguished by the sphere distances is much smaller.

Tracking in wide areas typically means establishing the **camera pose of a mobile device** relative to a stationary environment.

Wide-area tracking can be realized by instrumenting an area with fiducials on the wall, if visual pollution is not a concern.

Optical Tracking - Target Identification

Natural Feature Target Identification

Natural feature point recognition can scale to **hundreds of thousands or even millions of feature points**.

An **individual feature point recognition** in a query image is not discriminative enough for reliable identification of a location and the **co-occurrence of feature points** in an image becomes essential.

A feature can be **assumed matched only if a certain number of features** from points that are known to be close in the real world appear together in an image.

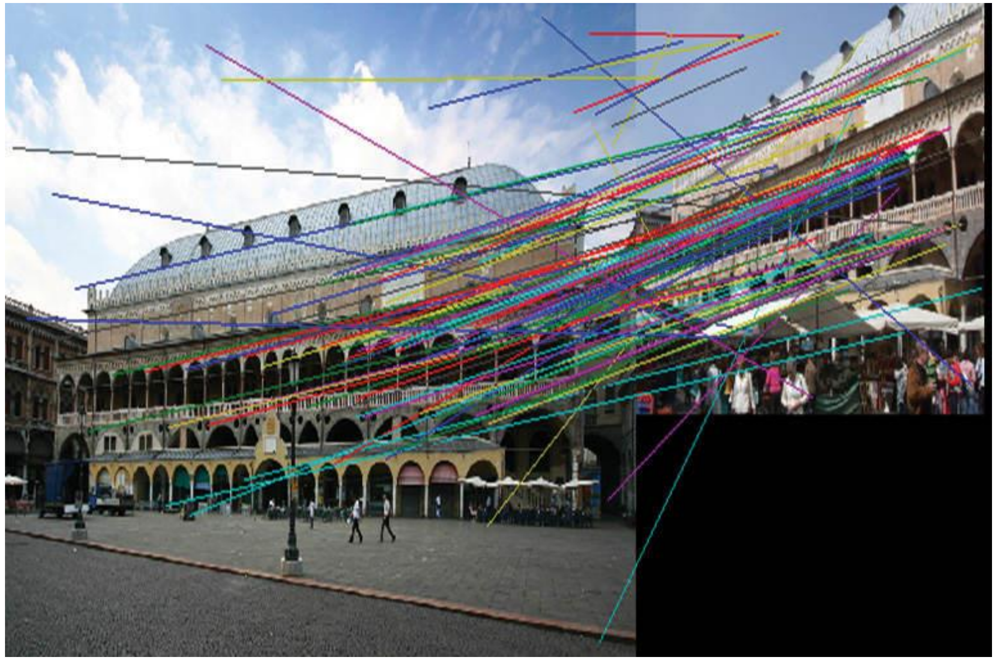
Optical Tracking - Target Identification

Natural Feature Target Identification

Tens or hundreds of feature point candidates extracted from a single image are matched with the massive feature point database.

Robust statistical techniques are used to exclude spurious matches and determine the most likely matches among the inliers.

From enough correct point associations, a scene can be identified, and the current camera pose can be determined.



Sensor Fusion

AR systems usually have **several sensors**, optical and non-optical.

Best results are obtained when tracking makes use of the input provided by all available sensors (**sensor fusion**).

However, hybrid tracking system **increases the weight, cost,** and **power consumption** of the resulting system, and **requires calibration** between sensors.

Sensor Fusion - Complementary Sensor Fusion

Complementary sensor fusion occurs when multiple sensors supply **different degrees of freedom**.

The most common use of complementary sensor fusion is to combine a **position-only** sensor with an **orientation-only** sensor to yield full 6DOF.

For example, in a modern mobile phone, **GPS delivers position** information, while the **compass and accelerometer deliver orientation** data.

Sensor Fusion - Competitive Sensor Fusion

Competitive sensor fusion combines the data from different sensor types measuring the **same degree of freedom independently**.

The individual measurements are **combined into a measurement of superior quality** using some form of mathematical fusion, for example, Kalman filtering.

Sensor fusion is a good approach for combining slow and fast sensors, and absolute and relative sensors.

An example is an **inertial measurement unit** (IMU) for orientation measurement, where a full IMU configuration consists of **three orthogonal magnetometer, gyroscope, and accelerometer units** each.

Sensor Fusion - Cooperative Sensor Fusion

In **cooperative sensor fusion**, a primary sensor relies on information from a secondary sensor to obtain its measurements.

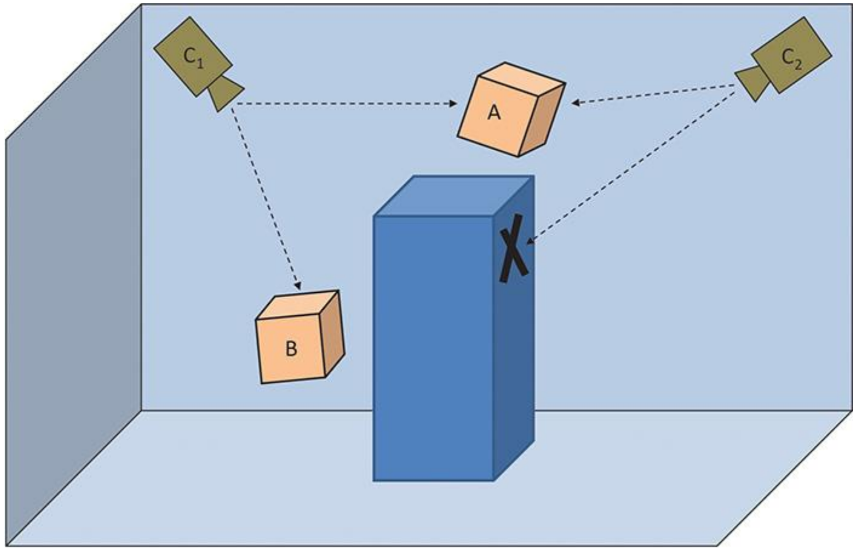
Cooperative sensor fusion can be described as any measurement of a property that cannot be derived from either sensor alone.

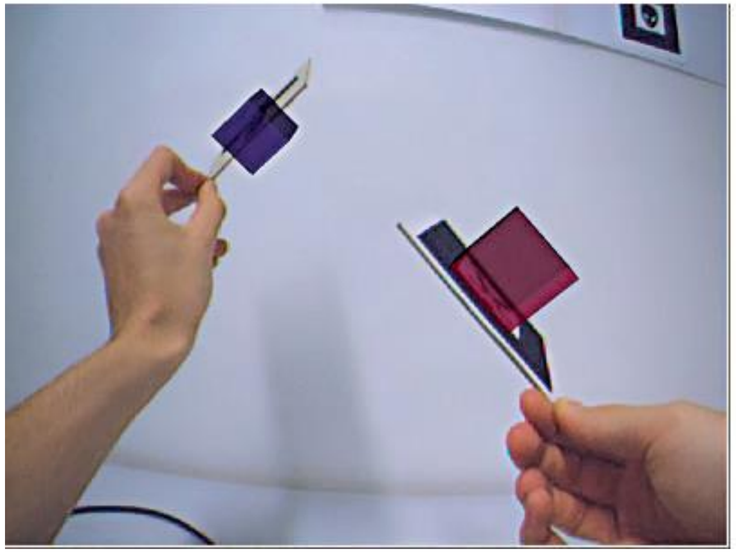
Stereo camera rigs used in optical tracking can be seen as performing cooperative sensor fusion, since their known epipolar geometry allows converting two 2D measurements to a single 3D measurement (or RGB-D and 360º cameras).

Sensor Fusion - Cooperative Sensor Fusion

Another application of cooperative sensor fusion is the **combination of inside-out and outside-in** tracking—that is, a mobile sensor and a stationary sensor.

This kind of configuration can track target objects that are occluded from the stationary sensor's location.





Bibliography

Based on:

[1] - Augmented Reality: Principles and Practice, 1st Edition by Dieter Schmalstieg and Tobias Hollerer, June 2016, Pearson Education

[2] - Practical Augmented Reality: A Guide to the Technologies, Applications, and Human Factors for AR and VR, Steve Aukstakalnis, 2017, Addison-Wesley