AA274A: Principles of Robot Autonomy I
Course Notes

Oct 15, 2019

8 Introduction to Robot Sensors

One of the most important tasks of an autonomous system of any kind is to acquire knowledge about its environment. This is done by taking measurements using various sensors and then extracting meaningful information from those measurements. In this lecture we present the most common sensors used in mobile robots and then discuss strategies for extracting information from the sensors. The last part of this lecture focuses on robot vision, which is one of the most powerful and thoroughly studied robot sensors. Most of this lecture note is a direct excerpt from [SNS11], in case more information is required.

8.1 Sensors for Mobile Robots

8.1.1 Sensor Classification

We classify sensors using two important functional axes: *proprioceptive/exteroceptive* and *passive/active*. *Proprioceptive* sensors measure values internal to the system (robot), for example, motor speed, wheel load, robot arm joint angles, and battery voltage. *Exteroceptive* sensors acquire information from the robot's environment, for example, distance measurements, light intensity, and sound amplitude. Hence exteroceptive sensor measurements are interpreted by the robot in order to extract meaningful environmental features.

8.1.2 Characterizing sensor performance

The sensors we describe in this chapter vary greatly in their performance characteristics. Some sensors provide extreme accuracy in well-controlled laboratory settings but are overcome with error when subjected to real-world environmental variations. Other sensors provide narrow, high-precision data in a wide variety of settings. In order to quantify such performance characteristics, first we formally define the sensor performance terminology that will be valuable for the rest of this quarter. There are largely two categories of sensor performance: *Basic sensor response ratings* that can be characterized as design specifications, and *in situ* performance that measures how well a sensor performs in the real environment.

Sensor response ratings. A number of sensor characteristics can be rated quantitatively in a laboratory setting. Such performance ratings will necessarily be best-case scenarios when the sensor is placed on a real-world robot, but are nevertheless useful. *Dynamic range* is used to measure ratio between the lower and upper limits of input values to the sensor under normal sensor operation. Because this raw ratio can be unwieldy, it is usually measured in decibels, which are computed as ten times the common logarithm of the dynamic range. Range is also an important rating in mobile robot applications because often robot sensors operate in environments where they are frequently exposed to input values beyond their working range. In such cases, it is critical to understand how the sensor will respond. For example, an optical rangefinder will have a minimum operating range and can thus provide spurious data when measurements are taken with the object closer than that minimum.

Resolution is the minimum difference between two values that can be detected by a sensor. Usually, the lower limit of the dynamic range of a sensor is equal to its resolution. However, in the case of digital sensors, this is not necessarily so.

Linearity is an important measure governing the behavior of the sensor's output signal as the input signal varies. A linear response indicates that if two inputs x and y result in the two outputs $f(x)$ and $f(y)$, then for any values a and b , $f(ax + by) = af(x) + bf(y)$. This means that a plot of the sensor's input/output response is simply a straight line.

Bandwidth or *frequency* is used to measure the speed with which a sensor can provide a stream of readings. Formally, the number of measurements per second is defined as the sensor's frequency in hertz. Because of the dynamics of moving through their environment, mobile robots often are limited in maximum speed by the bandwidth of their obstacle detection sensors. Thus, increasing the bandwidth of ranging and vision sensors has been a high-priority goal in the robotics community.

In situ sensor performance. These sensor characteristics can be reasonably measured in a laboratory environment with confident extrapolation to performance in real-world deployment. However, a number of important measures cannot be reliably acquired without deep understanding of the complex interaction between all environmental characteristics and the sensors in question. This is most relevant to the most sophisticated sensors, including active ranging sensors and visual interpretation sensors.

Sensitivity itself is a desirable trait. This is a measure of the degree to which an incremental change in the target input signal changes the output signal. Formally, sensitivity is the ratio of output change to input change. Unfortunately, however, the sensitivity of exteroceptive sensors is often confounded by undesirable sensitivity and performance coupling to other environmental parameters.

Cross-sensitivity is the technical term for sensitivity to environmental parameters that are orthogonal to the target parameters for the sensor. For example, a flux-gate compass can demonstrate high sensitivity to magnetic north and is therefore of use for mobile robot navigation. However, the compass will also demonstrate high sensitivity to ferrous building materials, so much so that its cross-sensitivity often makes the sensor useless in some indoor environments. High cross-sensitivity of a sensor is generally undesirable, especially when it cannot be modeled.

Error of a sensor is defined as the difference between the sensor's output measurements and the true values being measured, within some specific operating context. Given a true value v and a measured value m, we can define error as $e := m - v$.

Accuracy is defined as the degree of conformity between the sensor's measurement and the true value, and is often expressed as a proportion of the true value (e.g., 97.5% accuracy). Thus small error corresponds to high accuracy and vice versa: $a := 1 - |e|/v$. Of course, obtaining the ground truth, v , can be difficult or impossible, and so establishing a confident characterization of sensor accuracy can be problematic. Furthermore, it is important to distinguish between two different sources of error:

Systematic errors are caused by factors or processes that can in theory be modeled. These errors are, therefore, deterministic (i.e., predictable). Poor calibration of a laser rangefinder, an unmodeled slope of a hallway floor, and a bent stereo camera head due to an earlier collision are all possible causes of systematic sensor errors.

Random errors cannot be predicted using a sophisticated model; neither can they be mitigated by more precise sensor machinery. These errors can only be described in probabilistic terms (i.e., stochastically). Hue instability in a color camera, spurious rangefinding errors, and black level noise in a camera are all examples of random errors.

Precision is often confused with accuracy, and now we have the tools to clearly distinguish these two terms. Intuitively, high precision relates to reproducibility of the sensor results. For example, one sensor taking multiple readings of the same environmental state has high precision if it produces the same output. In another example, multiple copies of this sensor taking readings of the same environmental state have high precision if their outputs agree. Precision does not, however, have any bearing on the accuracy of the sensor's output with respect to the true value being measured. Suppose that the random error of a sensor is characterized by some mean value μ and a standard deviation σ . The formal definition of precision is the ratio of the sensor's output range to the standard deviation:

$$\text{precision} = \frac{\text{range}}{\sigma}. \tag{1}$$

Note that only σ and not μ has impact on precision. In contrast, mean error μ is directly proportional to overall sensor error and inversely proportional to sensor accuracy.

Characterizing error: The challenges in mobile robots. Mobile robots depend heavily on exteroceptive sensors. Many of these sensors concentrate on a central task for the robot: acquiring information on objects in the robot's immediate vicinity so that it may interpret the state of its surroundings. Of course, these "objects" surrounding the robot are all detected from the viewpoint of its local reference frame. Since the systems we study are mobile, their ever-changing position and their motion have a significant impact on overall sensor behavior. In this section, empowered with the terminology of the earlier discussions, we describe how dramatically the sensor error of a mobile robot disagrees with the ideal picture drawn in the previous section.

Blurring of systematic and random errors. Active ranging sensors tend to have failure modes that are triggered largely by specific relative positions of the sensor and environment targets. For example, a sonar sensor will produce specular reflections, producing grossly inaccurate measurements of range, at specific angles to a smooth sheetrock wall. During motion of the robot, such relative angles occur at stochastic intervals. This is especially true in a mobile robot outfitted with a ring of multiple sonars. The chances of one sonar entering this error mode during robot motion is high. From the perspective of the moving robot, the sonar measurement error is a random error in this case. Yet, if the robot were to stop, becoming motionless, then a very different error modality is possible. If the robot's static position causes a particular sonar to fail in this manner, the sonar will fail consistently and will tend to return precisely the same (and incorrect!) reading time after time. Once the robot is motionless, the error appears to be systematic and of high precision.

The fundamental mechanism at work here is the cross-sensitivity of mobile robot sensors to robot pose and robot-environment dynamics. The models for such cross-sensitivity are not, in an underlying sense, truly random. However, these physical interrelationships are rarely modeled, and therefore, from the point of view of an incomplete model, the errors appear random during motion and systematic when the robot is at rest.

Sonar is not the only sensor subject to this blurring of systematic and random error modality. Visual interpretation through the use of a CCD camera is also highly susceptible to robot motion and position because of camera dependence on lighting changes, lighting specularity (e.g., glare), and reflections. The important point is to realize that, while systematic error and random error are well defined in a controlled setting, the mobile robot can exhibit error characteristics that bridge the gap between deterministic and stochastic error mechanisms.

Multimodal error distributions. It is common to characterize the behavior of a sensor's random error in terms of a probability distribution over various output values. In general, one knows very little about the causes of random error, and therefore several simplifying assumptions are commonly used. For example, we can assume that the error is zero-mean in that it symmetrically generates both positive and negative measurement error. We can go even further and assume that the probability density curve is Gaussian. It is important for now to recognize the fact that one frequently assumes symmetry as well as unimodal distribution. This means that measuring the correct value is most probable, and any measurement that is farther away from the correct value is less likely than any measurement that is closer to the correct value. These are strong assumptions that enable powerful mathematical principles to be applied to mobile robot problems, but it is important to realize how wrong these assumptions usually are.

Consider, for example, the sonar sensor once again. When ranging an object that reflects the sound signal well, the sonar will exhibit high accuracy and will induce random error based on noise, for example, in the timing circuitry. This portion of its sensor behavior will exhibit error characteristics that are fairly symmetric and unimodal. However, when the sonar sensor is moving through an environment and is sometimes faced with materials that cause coherent reflection rather than return the sound signal to the sonar sensor, then the sonar will grossly overestimate the distance to the object. In such cases, the error will be biased toward positive measurement error and will be far from the correct value. The error is not strictly systematic, and so we are left modeling it as a probability distribution of random error. So the sonar sensor has two separate types of operational modes, one in which the signal does return and some random error is possible, and the second in which the signal returns after a multipath reflection and gross overestimation error occurs. The probability distribution could easily be at least bimodal in this case, and since overestimation is more common than underestimation, it will also be asymmetric.

As a second example, consider ranging via stereo vision. Once again, we can identify two modes of operation. If the stereo vision system correctly correlates two images, then the resulting random error will be caused by camera noise and will limit the measurement accuracy. But the stereo vision system can also correlate two images incorrectly, matching two fenceposts, for example, that are not the same post in the real world. In such a case stereo vision will exhibit gross measurement error, and one can easily imagine such behavior violating both the unimodal and the symmetric assumptions.

8.1.3 Representing uncertainty

So far, we've defined a terminology for describing the performance characteristics of a sensor. As mentioned there, sensors are imperfect devices with errors of both systematic and random nature. Random errors, in particular, cannot be corrected, and so they represent atomic levels of sensor uncertainty.

Statistical representation. We have already defined error as the difference between a sensor measurement and the true value. From a statistical point of view, we wish to characterize the error of a sensor, not for one specific measurement but for any measurement. Let us formulate the problem of sensing as an estimation problem. The sensor has taken a set of n measurements with values ρ_i. The goal is to characterize the estimate of the true value $E[X]$ given n measurements with values ρ_i, and is:

$$E[X] = g(\rho_1, \rho_2, ..., \rho_n).$$

The probability density function $f(x)$ is used to characterize the statistical properties of X. The sum of all probabilities is:

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

The expected value $E[X]$ is represented by the mean:

$$\mu = E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

and the characterization of the range of possible values is defined as the variance:

$$Var(X) = \sigma^2 = \int_{\infty}^{\infty} (x - \mu)^2 f(x) dx$$

where the standard deviation is σ.

Independence of random variables. With the tools presented here, we often evaluate systems with multiple random variables. For instance, a mobile robot's laser rangefinder may be used to measure the position of a feature on the robot's right and, later, another feature on the robot's left. The position of each feature in the real world may be treated as random variables, X_1 and X_2. Two random variables X_1 and X_2 are independent if the particular value of one has no bearing on the particular value of the other. In this case we can draw several important conclusions about the statistical behavior of X_1 and X_2. First, the expected value (or mean value) of the product of random variables is equal to the product of their mean values:

$$E[X_1 X_2] = E[X_1]E[X_2].$$

Second, the variance of their sums is equal to the sum of their variances:

$$Var(X_1 + X_2) = Var(X_1) + Var(X_2).$$

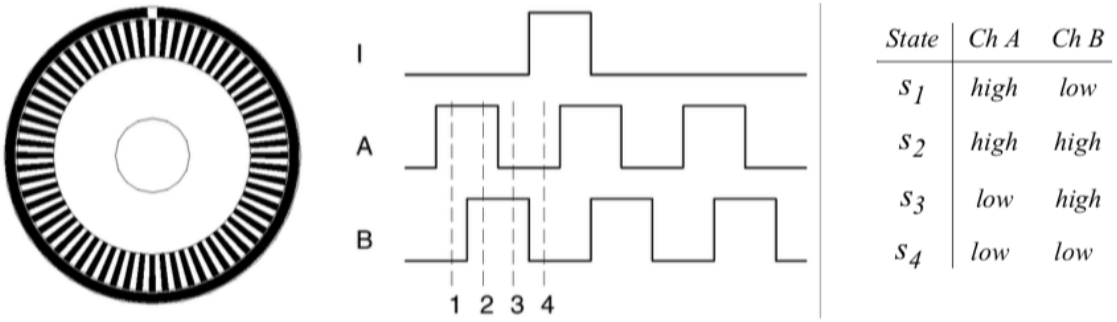
In mobile robotics, we often assume the independence of random variables even when this assumption is not strictly true. The simplification that results makes a number of the existing mobile robot-mapping and navigation algorithms tenable, as described in later lectures on localization. A further simplification, described in the following section, revolves around one specific probability density function used more often than any other when modeling error: the Gaussian distribution.

Gaussian distribution. The Gaussian distribution, also called the *normal distribution*, is used across engineering disciplines when a well-behaved error model is required for a random variable for which no error model of greater felicity has been discovered. The Gaussian has many characteristics that make it mathematically advantageous to other ad hoc probability density functions. It is symmetric around the mean μ. There is no particular bias for being larger than or smaller than μ, and this makes sense when there is no information to the contrary. The Gaussian distribution is also unimodal, with a single peak that reaches a maximum at μ (necessary for any symmetric, unimodal distribution). This distribution also has tails (the value of $f(x)$ as x approaches $-\infty$ and ∞) that approach zero only asymptotically. This means that all amounts of error are possible, although very large errors may be highly improbable. In this sense, the Gaussian is conservative. Finally, as seen in the formula for the Gaussian probability density function, the distribution depends on only two parameters μ and σ, so the distribution can be written as:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{\sigma^2} \right).$$

8.1.4 Common Sensors on Mobile Robots.

Wheel/motor sensors. Wheel/motor sensors are devices used to measure the internal state and dynamics of a mobile robot. These sensors have vast applications outside of mobile robotics and, as a result, mobile robotics has enjoyed the benefits of high-quality, low-cost wheel and motor sensors that offer excellent resolution. In mobile robotics, encoders are one of the most popular means to control the position or speed of wheels and other motor-driven joints. Because these sensors are proprioceptive, their estimate of position is best in the reference frame of the robot and, when applied to the problem of robot localization, significant corrections are required.



State	*Ch A*	*Ch B*
s_1	*high*	*low*
s_2	*high*	*high*
s_3	*low*	*high*
s_4	*low*	*low*

Figure 1: Quadrature optical wheel encoder [SNS11]

An optical encoder is basically a mechanical light chopper that produces a certain number of sine or square wave pulses for each shaft revolution. It consists of an illumination source, a fixed grating that masks the light, a rotor disc with a fine optical grid that rotates with the shaft, and fixed optical detectors. As the rotor moves, the amount of light striking the optical detectors varies based on the alignment of the fixed and moving gratings. In robotics, the resulting sine wave is transformed into a discrete square wave using a threshold to choose between light and dark states. Resolution is measured in cycles per revolution (CPR). The minimum angular resolution can be readily computed from an encoder's CPR rating. A typical encoder in mobile robotics may have 2000 CPR, while the optical encoder industry can readily manufacture encoders with 10,000 CPR. In terms of required bandwidth, it is of course critical that the encoder be sufficiently fast to count at the shaft spin speeds that are expected. Industrial optical encoders present no bandwidth limitation to mobile robot applications.

Usually in mobile robotics the quadrature encoder is used. In this case, a second illumination and detector pair is placed 90 degrees shifted with respect to the original in terms of the rotor disc. The resulting twin square waves, shown in Figure 1, provide significantly more information. The ordering of which square wave produces a rising edge first identifies the direction of rotation. Furthermore, the four detectably different states improve the resolution by a factor of four with no change to the rotor disc. Thus, a 2000 CPR encoder in quadrature yields 8000 counts.

As with most proprioceptive sensors, encoders are generally in the controlled environment of a mobile robot's internal structure, and so systematic error and cross-sensitivity can be engineered away. The accuracy of optical encoders is often assumed to be 100% and, although this may not be entirely correct, any errors at the level of an optical encoder are dwarfed by errors downstream of the motor shaft.

Heading sensors. Heading sensors can be proprioceptive (gyroscope, inclinometer) or exteroceptive (compass). They are used to determine the robot's orientation and inclination. They allow us, together with appropriate velocity information, to integrate the movement to a position estimate. This procedure, which has its roots in vessel and ship navigation, is called dead reckoning.

Compasses. Compasses are an example of exteroceptive heading sensors. Digital compasses using the Hall effect are popular in mobile robotics. Using the earth's magnetic field, they provide a rough estimate of direction. They are inexpensive, but often suffer from poor resolution and accuracy. Flux gate compasses have improved resolution and accuracy, but come at a larger price and physical size. Both compass types are vulnerable to vibrations and disturbances in the magnetic field, and are therefore less well suited for indoor applications.

Gyroscopes. Gyroscopes are heading sensors that preserve their orientation in relation to a fixed reference frame. Thus, they provide an absolute measure for the heading of a mobile system. Gyroscopes can be classified in two categories, mechanical gyroscopes and optical gyroscopes.

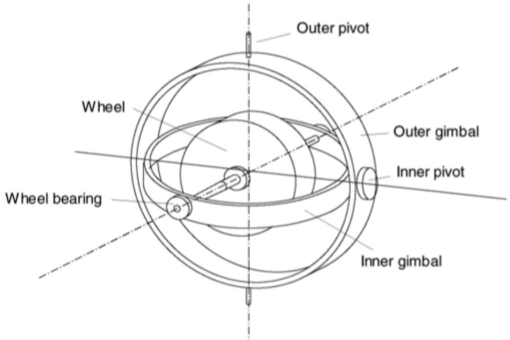


Figure 2: Two-axis mechanical gyroscope [SNS11]

Mechanical gyroscopes. The concept of a mechanical gyroscope relies on the inertial properties of a fast-spinning rotor. The property of interest is known as the gyroscopic precession. If you try to rotate a fast-spinning wheel around its vertical axis, you will feel a harsh reaction in the horizontal axis. This is due to the angular momentum associated with a spinning wheel and will keep the axis of the gyroscope inertially stable. The reactive torque τ and thus the tracking stability with the inertial frame are proportional to the spinning speed ω, the precession speed Ω, and the wheel's inertia I.

$$\tau = I\omega\Omega$$

By arranging a spinning wheel, as seen in Figure 2, no torque can be transmitted from the outer pivot to the wheel axis. The spinning axis will therefore be space-stable (i.e., fixed in an inertial reference frame). Nevertheless, the remaining friction in the bearings of the gyro axis introduce small torques, thus limiting the long-term space stability and introducing small errors over time. A high quality mechanical gyroscope can cost up to \$100,000 and has an angular drift of about 0.1 degrees in 6 hours.

Optical gyroscopes are a relatively new invention. They use angular speed sensors with two monochromatic light beams, or lasers, emitted from the same source. Two beams are sent clock - and counterclockwise through an optical fiber. Since the laser traveling in the direction of rotation has a slightly shorter path, it will have a higher frequency. This frequency difference δf is proportional to the angular velocity, which can therefore be estimated. In modern optical gyroscopes, bandwidth can easily exceed 100 kHz, while resolution can be smaller than 0.0001 degrees/hr.

Accelerometer. An accelerometer is a device used to measure all external forces acting upon it, including gravity. Essentially it is a simple spring-mass-damper system that can be represented by this second order differential equation [DJ08]:

$$F_{applied} = F_{inertial} + F_{damping} + F_{spring} = m\ddot{x} + c\dot{x} + kx$$

where m is the proof mass, c is the damping coefficient, k is the spring constant, and x is the equilibrium case relative position. When a static force is applied, the system will oscillate until it reaches a steady state. Appropriate damping material and mass are chosen to ensure that the system can stabilize quickly and then the applied acceleration can be calculated as

$$a_{applied} = \frac{kx}{m}.$$

Modern accelerometers, like the ones in mobile phones, are usually very small, which is enabled by the Micro Electro-Mechanical Systems (MEMS) consisting of a cantilevered beam and a proof mass. The deflection of the proof mass from its neutral position is measured using the capacitive effect, which essentially measures the capacitance, or the piezoelectric effect, which measures an induced voltage.

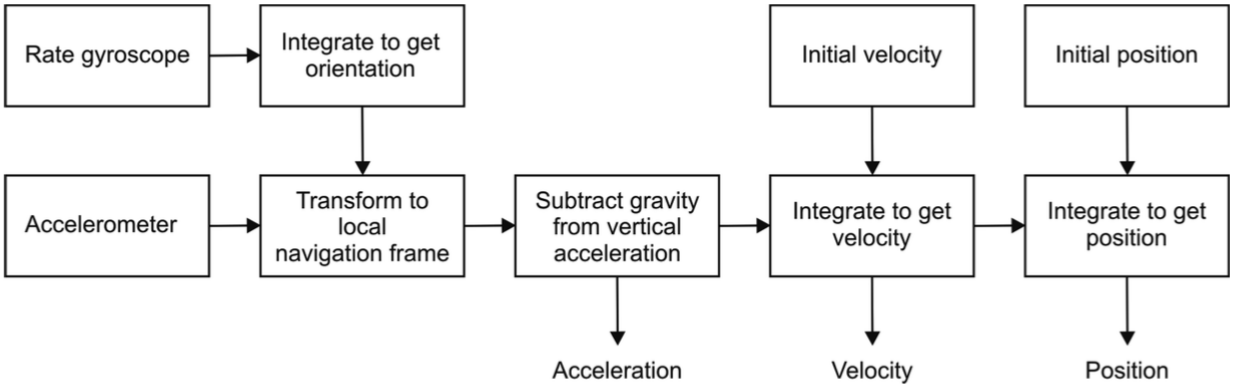
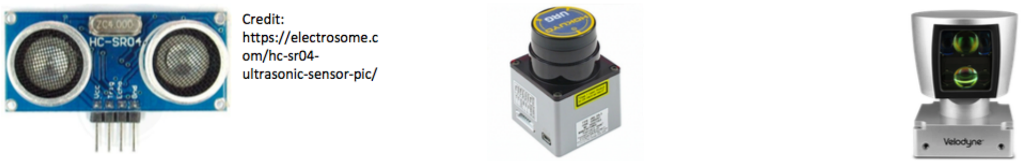

Rate gyroscope
Integrate to get orientation
Initial velocity
Initial position
Accelerometer
Transform to local navigation frame
Subtract gravity from vertical acceleration
Integrate to get velocity
Integrate to get position
Acceleration
Velocity
Position

Figure 3: IMU block diagram [DJ08]

Inertial measurement unit (IMU). An inertial measurement unit (IMU) is a device that uses gyroscopes and accelerometers to estimate the relative position, velocity, and acceleration of a moving vehicle. An IMU estimates the six-degree-of-freedom (DOF) pose of the vehicle: position (x, y, z) and orientation (roll, pitch, yaw). Nevertheless, heading sensors like compasses and gyroscopes, which conversely only estimate orientation, are often improperly called IMUs.

Besides the 6-DOF pose of the vehicle, commercial IMUs also usually estimate velocity and acceleration. To estimate the velocity, the initial speed of the vehicle needs to be known. The working principle of an IMU is shown in Figure 3. Let us suppose that our IMU has three orthogonal accelerometers and three orthogonal gyroscopes. The gyroscope data is integrated to estimate the vehicle orientation while the three accelerometers are used to estimate the instantaneous acceleration of the vehicle. The acceleration is then transformed to the local navigation frame by means of the current estimate of the vehicle orientation relative to gravity. At this point the gravity vector can be subtracted from the measurement. The resulting acceleration is then integrated to obtain the velocity and then integrated again to obtain the position, provided that both the initial velocity and position are a priori known. To overcome the need of knowing of the initial velocity, the integration is typically started at rest (i.e., velocity equal to zero).

Observe that IMUs are extremely sensitive to measurement errors in both gyroscopes and accelerometers. For example, drift in the gyroscope unavoidably undermines the estimation of the vehicle orientation relative to gravity, which results in incorrect cancellation of the gravity vector. Additionally observe that, because the accelerometer data is integrated twice to obtain the position, any residual gravity vector results in a quadratic error in position. Because of this and the fact that any other error is integrated over time, drift is a fundamental problem in IMUs. After long period of operation, all IMUs drift. To cancel this drift, some reference to some external measurement is required. In many robot applications, this has been done using cameras or GPS.



Credit: https://electrosome.com/hc-sr04-ultrasonic-sensor-pic/

Figure 4: (A) Ultrasonic sensor; (B) laser rangefinder; (C) laser range sensor

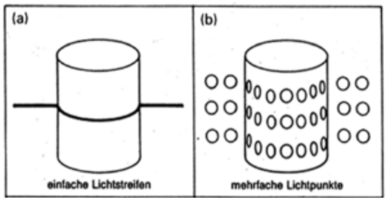


Figure 5: Possible light structures [SNS11]

Ground Beacons. Beacons are signaling devices with precisely known positions. An example from intuition is the lighthouse. Position can be determined by knowing the position of the beacon. More advanced examples of beacons are GPS, motion capture system for indoor use. With at least twenty-four operational GPS satellites at all times, the GPS synchronizes and reads data transmission from four satellites to obtain its own position based on the arrival time and instantaneous location. The four satellites provide three position axes plus a time correction. The localization resolution of GPS can be achieved using pseudorange with an extension method called differential GPS that uses a second receiver that is static at a known exact position to correct errors.

Active Ranging. Active ranging sensors provide direct measurements of distance to objects in vicinity. These sensors are important for both localization and environment reconstruction. For example, they can be used for self-driving cars to know where the destinations are. There are two types of active ranging sensors. One is the time-of-flight active ranging sensors (e.g., ultrasonic, laser rangefinder, and time-of-flight camera) and the other is geometric active ranging sensors (optical triangulation and structured light).

Time-of-flight Active Ranging makes use of the propagation speed of sounds or an electromagnetic wave. The travel distance is given by

$$d = c \cdot t$$

where d is the distance traveled, c is the speed of wave propagation, and t is the time of flight. It is important to point out that the propagation speed v of sound is approximately $0.3m/ms$ whereas the speed of electromagnetic signals is $0.3m/ns$, which is 1 million times faster. The time of flight for a typical distance, say $3m$, is $10ms$ for an ultrasonic system but only $10ns$ for a laser rangefinder. It is thus evident that measuring the time of flight t with electromagnetic signals is more technologically challenging. This explains why laser range sensors have only recently become affordable and robust for use on mobile robots. The quality of time-of-flight range sensors depends mainly on:

- uncertainties in determining the exact time of arrival of the reflected signal;
- inaccuracies in the time-of-flight measurement (particularly with laser range sensors);
- the dispersal cone of the transmitted beam (mainly with ultrasonic range sensors);
- interaction with the target (e.g., surface absorption, specular reflections);
- variation of propagation speed;
- the speed of the mobile robot and target (in the case of a dynamic target)

As discussed in the following, each type of time-of-flight sensor is sensitive to a particular subset of this list of factors.

Geometric Active Ranging uses geometric properties in the measurements to establish distance readings. Optical triangulation sensors (1D) transmit a collimated beam toward the target and use lens to collect reflected light and project it onto a position-sensitive device or linear camera. Structured light sensors (2D or 3D) project a known light pattern (e.g., point, line, or texture) onto the environment. The reflection is captured by a receiver and then, together with known geometric values, range is estimated via triangulation. Figure 5 shows systems that project light textures. CCD or CMOS receiver can then take photos and filter these images to identify the pattern's reflection based on the geometrical deformation of the pattern of light.

Other Sensors. Some classical examples of other sensors include radars that use Doppler effect to produce velocity data. Tactile sensors are critical to mobile, autonomous robots and are well understood and easily implemented. Some emerging technologies include artificial skins for obtaining tactile measurements and neuromorphic camera that detect motions with neurons spiking in changes of illumination. For example, small drones are often equipped with neuromorphic cameras which have a very small power consumption.

Vision sensors. Vision is our most powerful sense. It provides us with an enormous amount of information about the environment and enables rich, intelligent interaction in dynamic environments. It is therefore not surprising that a great deal of effort has been devoted to providing machines with sensors that mimic the capabilities of the human vision system. The first step in this process is the creation of sensing devices that capture the light and convert it into a digital image. The second step is the processing of the digital image in order to get salient information like depth computation, motion detection, color tracking, feature detection, scene recognition, and so on. Because vision sensors have become very popular in robotic applications, the remaining lecture will be dedicated to the fundamentals of computer vision and image processing and their use in robotics.

8.2 Fundamentals of Computer Vision

The analysis of images and their processing are two major fields that are known as computer vision and image processing. The years between 1980 and 2019 have seen tremendous advances and new theoretical findings in these fields and some of the most sophisticated computer vision and image processing techniques have found many industrial applications in consumer cameras, photography, defect inspection, monitoring and surveillance, video games, movies, and the like.

The remaining parts of this lecture are dedicated to these two fields. First, we will introduce the working principle of the camera and the image formation. Starting next lecture, we will present two ways of estimating the depth, which are depth from focus and stereo vision.

8.2.1 The digital camera

While the basic idea of a camera has existed for thousands of years, the first clear description of one was given by Leonardo Da Vinci in 1502 and the oldest known published drawing of a *camera obscura*, a dark room with a pinhole to image a scene, was shown by Gemma Frisius in 1544. By 1685, Johann Zahn had designed the first portable camera, and in 1822, Joseph Nicephore Niepce took the first physical photograph.

A modern camera in general can be defined as a sensor that captures light and converts that signal into a digital image. Light falling on an imaging sensor is usually picked up by an active sensing area, integrated for the duration of the exposure (usually expressed as the shutter speed, e.g., $1/125, 1/60, 1/30$ of a second), and then passed to a set of sense amplifiers. The two main kinds of ensors used in digital still and video cameras today are CCD (charge coupled device) and CMOS (complementary metal oxide on silicon).

The CCD chip is an array of light-sensitive picture elements, or pixels, usually with between 20,000 and several million pixels total. Each pixel can be thought of as a light-sensitive, discharging capacitor that is 5 to $25\mu m$ in size. The complementary metal oxide semiconductor (CMOS) chip is a significant departure from the CCD. It, too, has an array of pixels, but located along the side of each pixel are several transistors specific to that pixel.Traditionally, CCD sensors outperformed CMOS in quality sensitive applications such as digital single-lens-reflex cameras, while CMOS was better for low-power applications, but today, CMOS is used in most digital cameras.

8.2.2 Image formation

Before reaching the photon sensor in cameras, light rays first originate from a light source, which emits them in various wavelengths and directions. Averaged over time, the emitted wavelengths and directions can be precisely described using a probability distribution function specific to the characteristics of the light source. When the light rays strike an object, they either directly reflect, scatter, or are absorbed depending on the surface properties of

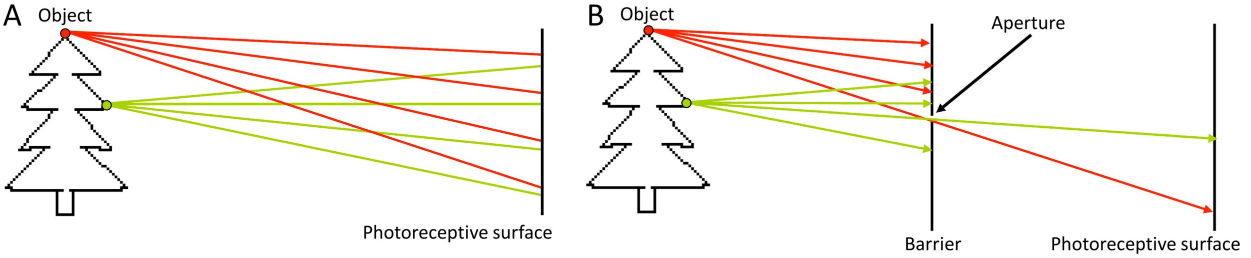


Figure 6: Light rays on a photoreceptive surface referred to as the image plane. In (A), the image is blurry whereas in (B), a barrier has been added so that ”red” and ”green” scattered light rays can be distinguished.

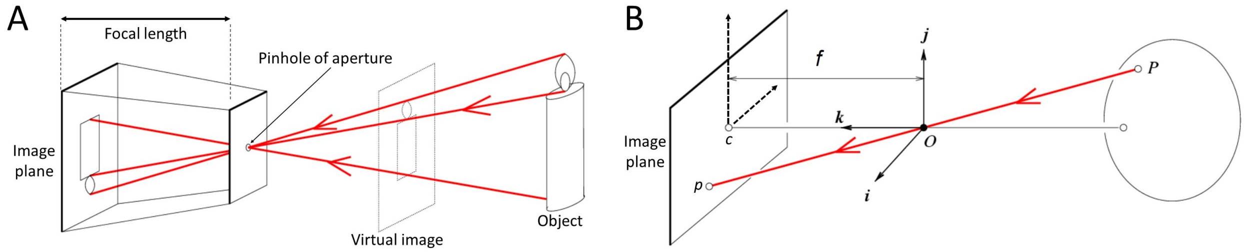


Figure 7: Pinhole camera model. Due to the geometry of the pinhole camera system, the object image is inverted on the image plane in (A). In (B), O is the camera center and p the principal point. The camera center is here placed at the coordinate origin. Note the image plane is placed in front of the camera center.

the object and the wavelength of the light. For example, an object looks blue because blue wavelengths of light are primarily scattered off the surface while other wavelengths are absorbed, a black object looks black because it absorbs most of the light rays, and a perfect mirror reflects all visible wavelengths of light.

Cameras work by capturing these light rays on photoreceptive surface, most often a CCD or a CMOS sensor. But the light rays must pass through the lens before reaching the sensor. As we see in Figure 6A, if we simply place a planar photoreceptive surface, or *image plane*, in front of an object, light rays that scattered from multiple different locations on the object will arrive at similar locations on the image plane and only an extremely blurred image of the object will be recorded. A solution to this blurring issue, as seen in Figure 6B, is to add a barrier that blocks off most of the light and only lets light through an aperture. The earliest means of filtering these light rays was done by a small hole on the flat surface, which led to the invention of the first example of camera in the history, the pinhole camera (or *camera obscura*).

Pinhole camera model. A pinhole camera has no lens, but a single very small aperture. In short, it is a lightproof box with a small hole in one side. Light from the scene passes through this single point and projects an inverted image on the opposite side of the box

(Figure 7a). The importance of the pinhole camera is that its principle has also been adopted as a standard model for perspective cameras. This model can be derived directly from very simple thin lens model. When using the pinhole camera model, it is very important to remember that the pinhole corresponds to the center of the lens. This point is also commonly called *center of projection* of *optical center* (indicated with c in figure 7). The axis perpendicular to the *image plane*, which passes through the center of projection is called *optical axis*.

Perpective projection. To describe analytically the perspective projection operated by the camera, we have to introduce some opportune reference system wherein we can express the 3D coordinates of the scene point P and the coordinates of its projection p on the image plane. We will first consider a simplified model and finally the general model.

Let (X, Y, Z) be the camera reference frame with origin in C and Z-axis coincident with the optical axis. Assume also that the camera reference frame coincides with the world reference frame. This implies that the coordinates of the scene point P are already expressed in the camera frame. Let us also introduce a two-dimensional reference frame (i, j, k) for the image plane with origin in O and the i and j axes aligned as X and Y respectively as shown in figure 7. Finally, let $P = (X, Y, Z)$ and $p = (x, y)$. By means of simple considerations on the similarity of triangles, we can write:

$$\frac{f}{Z} = \frac{x}{X} = \frac{y}{Y}, \tag{2}$$

or equivalently,

$$\frac{x}{f} = \frac{X}{Z} \quad ; \quad \frac{y}{f} = \frac{Y}{Z} \tag{3}$$

We can then solve for x and y which lie on the image plane and are proportional to pixel coordinates which we will later discuss:

$$x = f\frac{X}{Z} \quad ; \quad y = f\frac{Y}{Z}$$

8.2.3 Camera lenses

While pinhole camera models are applicable to most modern cameras, there is one significant difference: a pinhole has a fixed aperture, or the opening on the barrier. In optics, a larger aperture means a greater number of light rays can pass through the aperture, which leads to blurring of the image. On the other hand, a smaller aperture means fewer light rays can pass through the aperture, which makes the resulting image darker. Depending on the environment, cameras need to adjust the size of aperture and therefore the amount of light. For example, the aperture can be very small when an image is taken outside under the bright sun, whereas more light needs to come through if the image is taken inside a very dark room. Lenses are the optical solution to a fixed hole on the wall: this optical element focuses light by means of refraction.

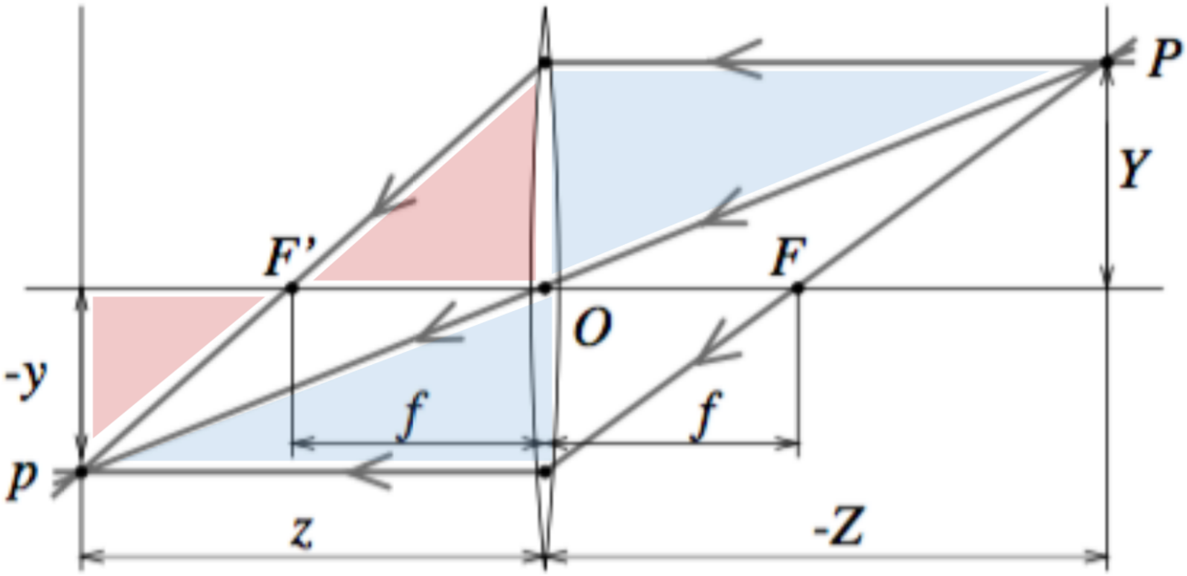


Figure 8: Depiction of the thin lens model.

By introducing a lens in our optical model, we can now extend our pinhole mathematical model. Figure 8 shows a diagram of the most basic lens model, which is the thin lens (i.e., no optical distortion due to the curvature of the lens). Snell's law states that rays passing through the center of the lens are not refracted, and those that are parallel to the optical axis are focused on the focal point labeled F'. In addition, all rays passing through P are focused by the thin lens on the point p. By inducing similar triangles, we can write the following equations first using the blue similar triangles, we obtain:

$$\frac{y}{Y} = \frac{z}{Z} \tag{4}$$

then the red similar triangles:

$$\frac{y}{Y} = \frac{z - f}{f} = \frac{z}{f} - 1.$$

Combining these two equations we can write the thin lens equation:

$$\frac{1}{z} + \frac{1}{Z} = \frac{1}{f}. \tag{5}$$

where f is the focal length. As you can see, this formula can also be used to estimate the distance to an object by knowing the focal length and the current distance of the image plane to the lens. This technique is called *depth from focus*. We can additionally notice from this geometry that the variable z in the lens model is equivalent to f in the pinhole model and once again notice that points that are not a distance of Z away from the lens will be out of focus. We also see that if Z approaches infinity, light would focus a distance of f away from the lens. Thus, we could assume a pinhole model if the lens is focused at infinity and this could allow us to map 3D points into the camera image plane where they could be converted into pixel coordinates.

8.2.4 Tradeoffs to optimize light and blur

With this more complete camera model, we can discuss the tradeoffs involved in most camera applications, especially those involving quickly-moving objects, i.e. high-speed cameras. Assuming we want a crisp image without blur, the following factors are critical: object movement, exposure time, receptivity of the photoreceptive surface, aperture size, and lighting conditions. For high speed applications, the exposure time of the camera, or amount of time that the photoreceptive imaging surface absorbs light, needs to be short so that the quickly moving object of interest doesn't move significantly during the exposure time. But with a lower exposure time, less light is absorbed. So to provide more light, we could open the aperture to a larger diameter, but doing so will decrease the depth of field. One solution is to increase the brightness of the light source or sources which is often why extremely bright light sources are used for high speed applications. Regardless, this fundamental issue of light capture exists for cameras especially high-speed cameras, and a proper balance of aperture size and exposure time is needed to optimize blur and image brightness. This, combined with a highly receptive photoreceptive surface and bright lighting conditions need to be considered carefully in photography.

References

[DJ08] Gregory Dudek and Michael Jenkin. *Inertial Sensors, GPS, and Odometry*, pages 477–490. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[SNS11] Roland Siegwart, Illah Reza Nourbakhsh, and Davide Scaramuzza. *Introduction to autonomous mobile robots.* MIT press, 2011.