AA 274: Principles of Robotic Autonomy Winter 2018

Lecture 4: Robotic sensors and introduction to computer vision

Scribes: Qizhan Tam, Ben Goeing, Bilan Jackie Yang, Marc Deetjen

4.1 Introduction

This lecture covers the performance characteristics of different types of sensors used in robots to detect their surroundings and introduces computer vision as used in robots. These lecture notes will begin by classifying different kinds of sensors and detailing their functionalities and will finish with an introduction to computer vision.

4.2 Robotic Sensors

4.2.1 Motivation and Sensor basics

To accomplish its objective, an autonomous system first requires knowledge about its state and environment. By using sensors, autonomous systems are able to extract useful information, which can then be used in subsequent decision-making. For example, self-driving cars use multiple sensors such as laser range finders (LiDARs) usually mounted on the top, stereo cameras for visual recognition, GPS to track its global position, radar for long range detection, and ultrasonic sensors for short range detection.

4.2.1.1 Classification of Sensors

Sensors can be divided in terms of the external or internal nature of their target measurement parameters. The first type, **proprioceptive sensors** measure the internal robot parameters. As example, wheel encoders are commonly used in mobile robots to detect angular speeds, which are then sent to the computer to make decisions about adjustments. Other examples include servo motors that measure the robot's joint angles and a battery voltage sensor to measure the battery level. The second type, **exteroceptive sensors** acquire information from the robot's environment. Examples include the sensors used in autonomous vehicles: LiDARs, radars, stereo cameras etc. These sensors can all used to detect the external environment such as the road markings and other cars.

Sensors can also be classified based on the interaction they have with their environment. **Passive sensors** measure ambient environmental energy entering the sensor. Temperature probes use electronic devices such as thermocouples that produce a temperature-dependent voltage. This process does not require any input energy from the user, it only requires the thermocouple to absorb or release heat energy to the environment until the temperature reaches equilibrium. In contrast, **active sensors** emit energy into the environment, and then measure the environmental reaction. As an example, radar sensors emit electromagnetic waves at specific frequencies. If the radar waves hit an object, they get reflected back to the receiver.The on-board computer of a car then registers the object and takes it into account when planning trajectories. However, emitted energy may affect the environment and sensors may suffer interference

from other signals. For example, if there are multiple neighboring cars using radar in an uncoordinated manner, they may blind each other from the radar interference caused.

4.2.1.2 Performance of Sensors

Design specifications are often used to characterize the performance of sensors. There are certain performance ratings that can be measured in a controlled, laboratory environment to a reasonable accuracy. **Dynamic Range** is the spread between the maximum and minimum input values to the sensor while maintaining normal sensor operation. For example, the sensors in our human body have a large dynamic range, even by the standards of modern electronic sensors. The human ear can detect a dynamic range of sound of roughly 140 dB [1]. It is also important to determine the level of detail that a sensor can provide. **Resolution** is the minimum difference between two values that can be detected by a sensor. For a sensor device, it may have multiple types of resolutions that are critical for its function. The usefulness of Magnetic Resonance Imaging (MRI) is largely due to the unprecedented temporal resolution of 20-30 ms and in-plane imaging resolution of 1.5-2.0mm, which can provide important information about diseases to health professionals [2]. Interpreting sensor signals can sometimes be challenging due to the behavior that sensors can exhibit under different operating conditions. For that, a measure of linearity can be useful. **Linearity** is the degree to which the sensor's output response depends linearly on the input. In other words, a perfectly linear sensor converts an input of some quantity into an output (usually electric signals), which is proportional to the input. This is preferred as it is mathematically easier to interpret the sensor outputs and to calibrate the sensor. **Bandwidth/frequency** is the speed with which a sensor can provide readings (in units of Hertz). This measure is important for mobile robots as their top speed is limited by the bandwidth of their obstacle detection sensors.

There are other performance ratings that must be tested in a real environment due to the complex interactions between the environment and the sensors, that are unable to be captured in a laboratory environment. One such performance rating is **sensitivity**. It is the degree to which a change in the target input changes the output signal. Sensitivity of a sensor to its target input is a generally good characteristic, however, a sensitive sensor may also be severely affected by non-target input signals that come from the environment. If a car's radar receiver has a high sensitivity to radar waves regardless of their source, it will result in severe interference in close proximity to other radar emitters. A performance rating closely related to sensitivity is **cross-sensitivity**. It is the sensitivity to environmental parameters that are unrelated to the target parameter. A flux-gate compass has a high sensitivity to the magnetic north but it is also very sensitive to ferrous objects, which renders it ineffective near such objects.

4.2.1.3 Sensor error analysis

In the real world, sensors have imperfections and can produce erroneous results. We therefore need to be able to characterize and predict sensor **errors**, which are defined as the difference between the sensor measurement m and its true value, v. $error := m - v$. We use **accuracy** to describe the conformity between the sensor's measurement and its true value. $accuracy := 1 - |error|/v$. While the average value of a sensor's measurements might agree well with the actual target value, there could be a large spread in the measurements. Therefore, **precision** is used to measure the reproducibility of the sensor results.

There are two main types of sensor errors that could occur. The first type, **systematic errors**, are caused by factors that can in theory be modeled which means they are deterministic. Such errors are usually a result of incorrect calibration. For example, forgetting to calibrate a simple weighing scale may result in a constant deviation from the correct weight. In contrast, **random errors** are errors that cannot be predicted

with sophisticated models which means they are stochastic, such as the spurious range-finding errors. To mitigate systematic and random errors, we often perform **error analysis** through probabilistic means. As simplifications, common assumptions include a symmetric, unimodal, Gaussian distribution which is convenient, but often a coarse simplification. We also use the error propagation law to predict how multiple errors from sensors could affect calculations of other parameters.

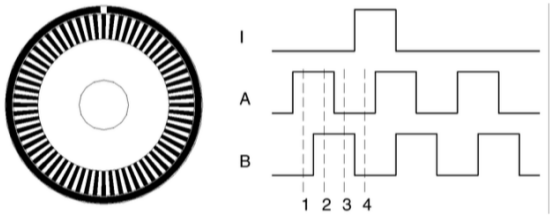
With the complexity of the suite of sensors that mobile robots have, it is not surprising that the basic termininologies and tools introduced in the above sections are still insufficient to cover the full range of difficulties in designing mobile robots that are robust to sensor errors. Instead of having neatly defined categories of errors, there could be the **blurring of systematic and random errors**. As an examples, active ranging sensors have failure modes caused by certain relative positions of the sensor with respect to environmental targets. In addition, **multimodal error distributions** could also occur. They represent the behavior of a sensor's random error in terms of a probability distribution over various output values, instead of the simplification that we previously introduced of probability distributions having symmetry and unimodal distributions. During normal operation with the waves reflecting off objects well, sonar will have errors with approximately symmetric and unimodal distributions. However, the specular reflections discussed before will cause the error probability distributions to be no longer symmetric nor unimodal.

4.2.2 Encoders

Encoders are proprioceptive sensors that can be used for tasks such as robot localization. In mobile robotics, optical encoders are used to measure angular speed and position and to control the position or speed of wheels and other motor-driven joints. This is best done in the reference frame of the robot.

Encoders are electro-mechanical devices that convert motion into a sequence of digital pulses. They consist of an illumination source, a fixed grating to mask the light, a rotor disc with a fine optical grid that rotates with the shaft, and fixed optical detectors. During the rotation of the motor, the amount of light that strikes the optical detectors varies based on the alignment of the fixed and moving gratings. This results in a sine wave, which is transformed into a discrete square wave using a given threshold to distinguish between "light" and "dark" states. The Resolution is then measured in cycles per revolution (CPR). Typical encoders have around 2000 CPR.

In mobile robotics, quadrature encoders are often used, in which case a second illumination and detector pair is added at a 90-degree angle. This results in twin square waves, which can provide additional information. (See Figure 4.1).



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

Figure 4.1: Quadrature optical wheel encoder [3]

4.2.3 Heading sensors

Heading sensors are used to determine the robots orientation and inclination. In combination with velocity information, they allow us to calculate a position estimate via integration. This process is referred to a dead reckoning.

4.2.3.1 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.

4.2.3.2 Gyroscopes

Gyroscopes are an example of proprioceptive heading sensors. They can provide an absolute measure for the heading of a mobile system. Mechanical gyroscopes contain an internal spinning wheel whose angular momentum keeps the axis of rotation inertially stable (see Figure 4.2). No torque τ can be transmitted from the outer frame onto the inner wheel, as it is proportional to the spinning speed ω, the precision speed Ω and the wheel's inertia I, as represented by the formula $\tau = I\omega\Omega$. High quality gyroscopes still have an angular drift of about 0.1 per 6h, due to friction.

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 the 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

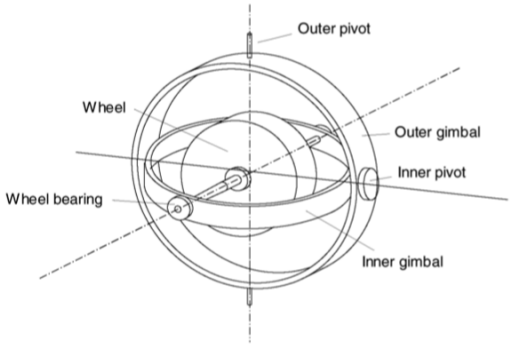


Figure 4.2: Two-axis mechanical gyroscope [3]

4.2.4 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 [4]:

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

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 then measured using either the capacitive effect or the piezoelectric effect.

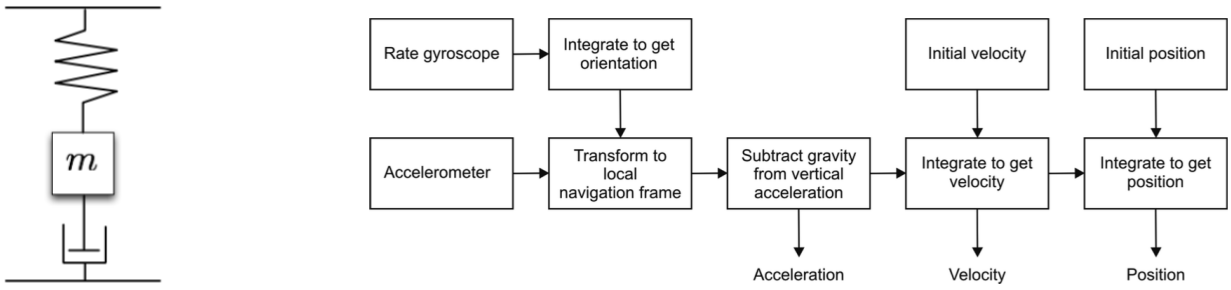


Figure 4.3: (A) Working principle of the mechanical accelerometer; (B) IMU block diagram [4]

4.2.5 Inertial measurement unit (IMU)

An IMU is a device that uses gyroscopes and accelerometers to estimate the relative position, orientation, velocity, and acceleration of a moving vehicle with respect to an inertial frame. It integrates the information from gyroscopes to estimate the vehicle orientation and from accelerometers to obtain the instantaneous acceleration to understand the position of the vehicle based on initial conditions. A detailed working principle is shown in Figure 4.3. Through multiple times of integration, however, noises are accumulated over time, which result in the fundamental problem of drifting in IMUs. To cancel drift, periodic references to external measurements are required. Thus, IMUs are usually augmented with other techniques like GPSs or cameras to obtain some reference to absolute positions.

4.2.6 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

Figure 4.4: (A) Ultrasonic sensor; (B) laser rangefinder; (C) laser range sensor

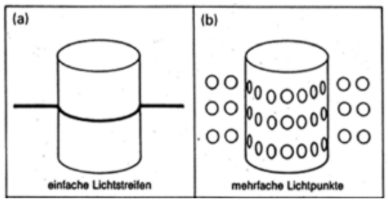


Figure 4.5: Possible light structures [3]

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.

4.2.7 Active Ranging

Active ranging sensors provide direct measurements of distance to objects in their 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 (using optical triangulation and structured light).

4.2.7.1 Time-of-flight Active Ranging

Time-of-flight active ranging sensors make 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. Commonly used active ranging sensors are the ultrasonic sensor and LiDAR. Distance can be obtained by measuring the time from the transmission of the waves to detection of an echo. These waves could be sound or light. Light sensors could be much more expensive because they are required to measure much shorter time but they also achieve much higher quality. LiDAR (Light Detection and Ranging) emits light beams and measures the time of return. Usually there are multiple emitters that receive multiple returns. LIDAR is often rotating to get 360 degree of view. The Performance depends on several factors, e.g., uncertainties in determining the exact time of arrival and interaction with the target.

4.2.7.2 Geometric Active Ranging

Geometric active ranging sensors use geometric properties in the measurements to establish distance readings. Optical triangulation sensors (1D) transmit a collimated beam toward the target and use a 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 4.5 shows systems that project light textures. CCD or CMOS receivers can then take photos and filter these images to identify the patterns reflection based on the geometrical deformation of the pattern of light.

4.2.8 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 cameras that can detect motions in neurons spiking in response to changes of illumination. For example, small drones are often equipped with neuromorphic cameras which also have very small power consumption.

4.3 Introduction to computer vision and cameras

Whereas most sensors require contact or an active signal to interact with their environment, vision sensors simply capture light rays that are already emitted by the environment. Vision is thus a powerful sensing method that can be defined as: the ability to interpret the surrounding environment using light in the visible spectrum reflected by objects in the environment. The most familiar vision sensor, the human eye, is an incredible example of a vision sensor because it captures an enormous amount of information on the order of millions of bits of information per second.

4.3.0.1 History of cameras

While the human eye has existed for thousands of years, the ancient idea of cameras has only been developed extensively more recently. Specifically, 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. Photography was born.

A modern camera in general can be defined as a sensor that captures light, converts that signal into a digital image, after which that image can be processed to filter desired information. These cameras capture images digitally by converting light into electric charge and processing it into electronic signals by one of two primary methods. Charge-Coupled Device (CCD) sensors transport charges from light rays across the chip so that it can be read into a voltage and recorded, whereas Complementary Metal-Oxide Semiconductor (CMOS) sensors use a transistor in each pixel combined with more traditional wires to record each pixel individually. Because CMOS sensors can be fabricated using traditional processes similar to microprocessor production, they are usually much cheaper than CCD sensors. Another benefit

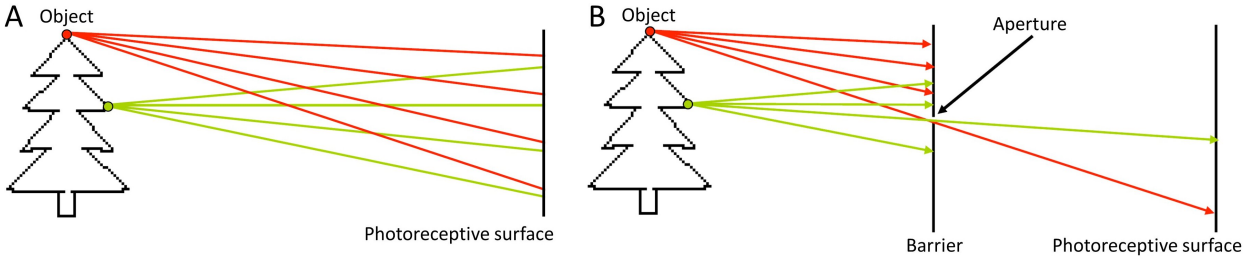


Figure 4.6: Light rays on a photoreceptive surface referred to as the image plane. In (A), the image is blurry whereas in (B), a barrier has been added so that "red" and "green" scattered light rays can be distinguished [5].

is that CMOS sensors consume much less power. However, CCD sensors usually have higher quality in terms of noise reduction and higher light sensitivity. Deciding which to use is application dependent.

4.3.0.2 Camera basics: the pinhole camera

The basic idea of a camera, meaning a sensor that captures light, involves an understanding of light rays. 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 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 some sort of photoreceptive surface in an organized way. As we see in Figure 4.6A, 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 4.6B, is to add a barrier that blocks off most of the light and only lets light through an aperture, or pinhole. Based on the geometry of this new system, if a light ray is absorbed in a certain location on the image plane, we know where that light ray last bounced off of an object. Namely, the light ray must have scattered from an object somewhere along the vector extending into space that connects the detected position on the image plane and the pinhole. In other words, the image plane will capture a picture of the object and surrounding scene because the geometry of the camera constrains the light rays that are captured. However, as captured by the photograph, the size of object is ambiguous and can only be determined if one knows how far away the object lies.

With this basic model for a camera, the geometry can be described with mathematical equations. First, from viewing Figure 4.7A, one can see that the object as recorded on the image plane is inverted, so often for mathematical derivations, the virtual image plane is used instead. Next, if we define an origin at the pinhole, o, in Figure 4.7B, we can say that point P on the object has coordinates $(-X, Y, -Z)$, while the point p on the image plane has coordinates $(x, -y, f)$ where f is the focal length. By drawing similar triangles, as the points P, o, and p are collinear, we can write the following equations where the negative signs cancel when using the virtual image plane:

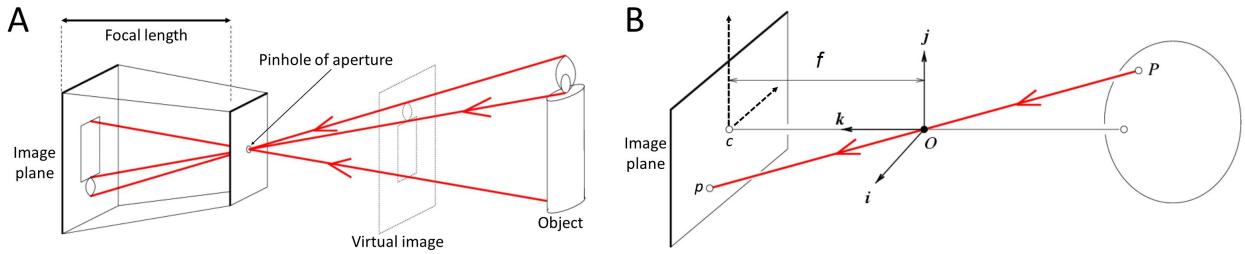


Figure 4.7: Pinhole camera model. In (A), it can be seen that due to the geometry of the pinhole camera system, the object image is inverted on the image plane and thus, for ease of math, the virtual image plane which displays the non-inverted object is used. In (B), the image center is labeled as c, the pinhole is labeled as o, the optical axis lies along vector k, the focal length is f, the 3D point is labeled as P, and the 2D pixel point is labeled p [6].

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

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

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

4.3.0.3 Camera lenses

One limitation of the current pinhole camera model is light. We have already discussed how lacking an aperture, i.e. having an aperture of infinite diameter, leads to a blurry image. But thus far in mathematical terms, we have technically been discussing an aperture of zero diameter which would in reality let in no light. In the middle ground between these two extremes, there is a tradeoff between amount of light that is absorbed by the photoreceptive surface and image blur. Additionally, for small aperture sizes, the resolution becomes fundamentally limited by diffraction in what is called a diffraction-limited system. One clever solution to the lighting-blur issue is to add a lens to replace the aperture as seen in Figure 4.8A. A lens is an optical element that focuses light by means of refraction. As can be seen however, the light from the object is only focused correctly if the distance to the object equals the focal length of the lens. For any other distance, the image will again appear blurred, although the closer to the focal length distance, the less blur. The size of the 3D region where the blur is acceptably low is called the depth of field.

With this lens, we can now extend our pinhole mathematical model. When looking at Figure 4.8B, we can use Snell's law and the assumption of a thin lens relative to the focal distance to develop similar triangles. First, rays that pass through the center of the lens are not refracted. Second, rays parallel to the optical axis are focused on the focal point labeled F'. Third, all rays passing through point P are focused by the thin lens on the point p. Thus, using similar triangles we can write the following equations first using the blue similar triangles then the red similar triangles:

$$\frac{y}{Y} = \frac{z}{Z} = \frac{z-f}{f}$$

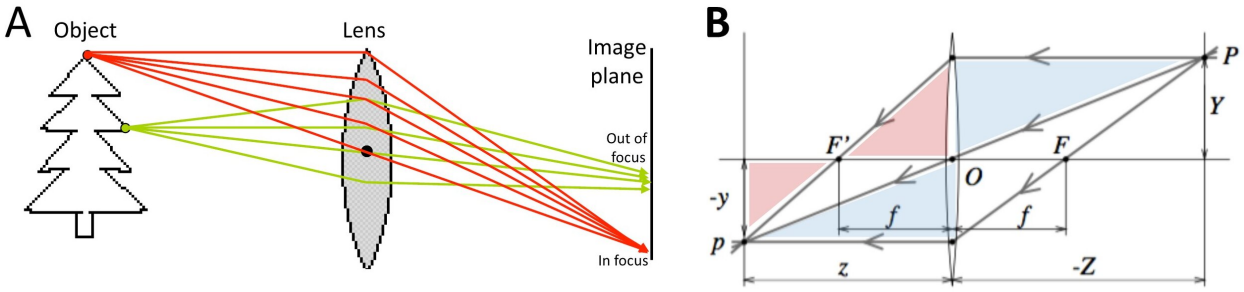


Figure 4.8: A lens is used to focus light. In (A), a lens is used to focus light originating from an object at a certain distance whereas different distances are blurred. In (B), the thin lens model can be developed using similar triangles [6].

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

$$\frac{1}{z} + \frac{1}{Z} = \frac{1}{f}$$

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.

4.3.0.4 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

[1] R. E. Runstein D. M. Huber. Dynamic range of human hearing. In *Modern Recording Techniques*. Academic Press, 2017.

[2] D Voit A Karaus KD Merboldt J Frahm M Uecker, S Zhang. Real-time mri at a resolution of 20 ms. *NMR Biomed*, 23:986–994, 2010a.

[3] D. Scaramuzza R. Siegwart, I. R. Nourbakhsh. Perception. In *Introduction to Autonomous Mobile Robots*. MIT Press, 2nd Edition, 2011.

[4] Gregory Dudek and Michael Jenkin. Inertial sensors, gps, and odometry. In *Springer Handbook of Robotics*. Springer, 2008.

[5] Jean Ponce David A. Forsyth. Geometric camera models. In *Computer Vision: A Modern Approach*. Prentice Hall, 2nd Edition, 2011.

[6] A. Zisserman R. Hartley. Camera models. In *Multiple View Geometry in Computer Vision*. Academic Press, 2002.