Perception – Sensors for mobile robots and Computer Vision I –

CSC398 Autonomous Robots

Ubbo Visser

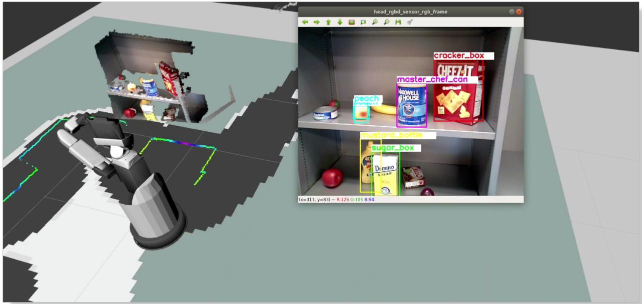
Department of Computer Science
University of Miami

November 14, 2024

UNIVERSITY
OF MIAMI

Outline

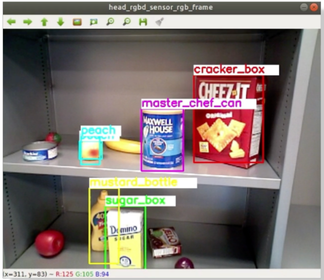
1. Sensor classification
2. Sensor performance
3. Computer Vision I



Perception - Sensors for mobile robots

Aim

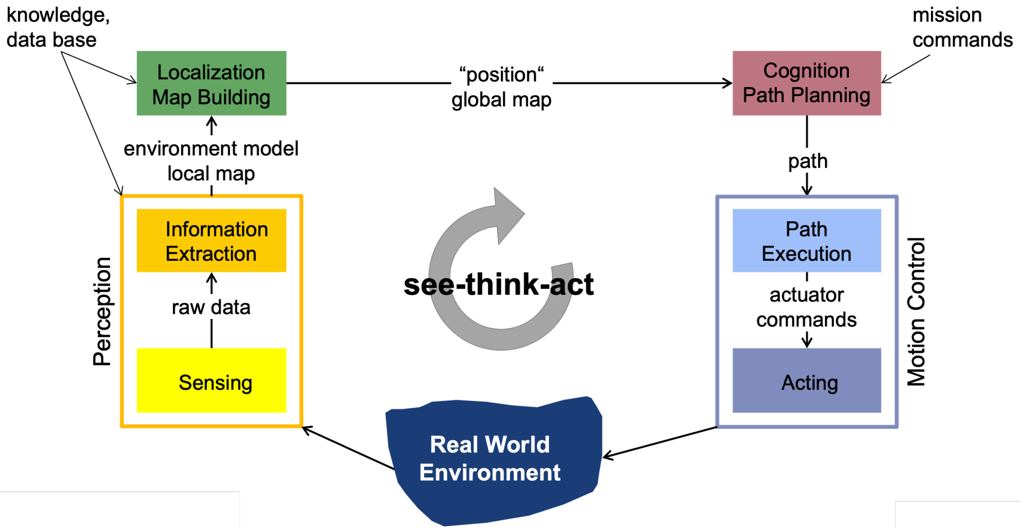
- Learn about key performance characteristics for robotic sensors, especially vision sensors
- Learn about a full spectrum of sensors, e.g. proprioceptive / exteroceptive, passive / active



Suggested Reading:

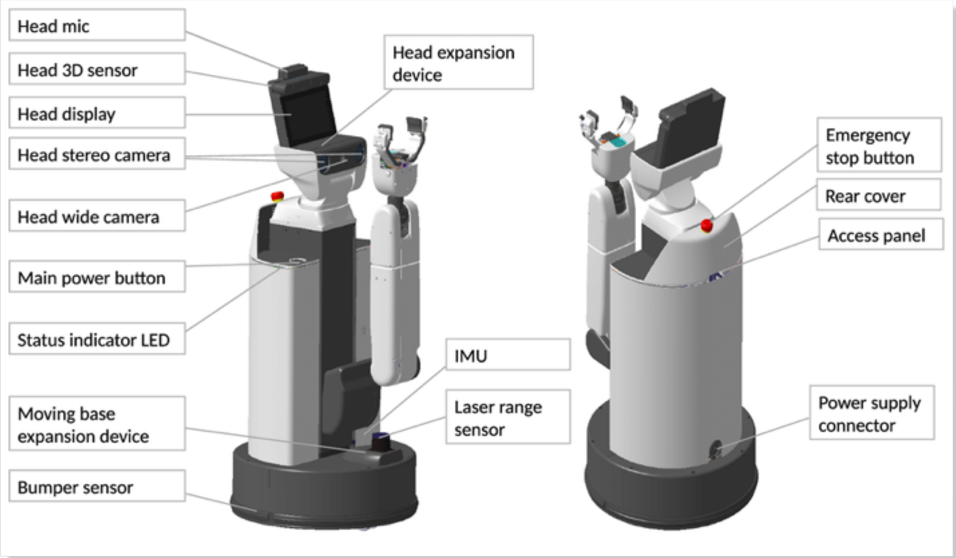
- *Introduction to Autonomous Mobile Robots* by Roland Siegwart, Illah Nourbakhsh, Davide Scaramuzza, The MIT Press, second edition 2011

Perception - Cognition - Action cycle



Source: Siegwart et. al (2018): Autonomous Mobile Robots, Lecture ETH Zürich

Example HSR


Head mic
Head 3D sensor
Head display
Head stereo camera
Head wide camera
Main power button
Status indicator LED
Moving base expansion device
Bumper sensor
Head expansion device
IMU
Laser range sensor
Emergency stop button
Rear cover
Access panel
Power supply connector

Sensor classification

- **Proprioceptive:** measure values internal to the robot, e.g.: motor speed, robot arm joint angles, and battery voltage
- **Exteroceptive:** acquire information from the robot's environment, e.g.: distance measurements and light intensity

- **Passive:** measure ambient environmental energy entering the sensor
  - Challenge: performance heavily depends on the environment
  - E.g.: temperature probes and cameras
- **Active:** emit energy into the environment and measure the reaction
  - Challenge: might affect the environment
  - E.g.: ultrasonic sensors and laser rangefinders

Basic sensor response ratings

- **Dynamic range:** ratio between the maximum and minimum input values (for normal sensor operation), usually measured in *decibels*
- **Resolution:** minimum difference between two values that can be detected by a sensor
- **Linearity:** whether the sensor's output response depends linearly on the input)
- **Bandwidth or frequency:** speed at which a sensor provides readings (in Hertz)

In situ sensor performance

- **Sensitivity:** ratio of output change to input change
- **Cross-sensitivity:** sensitivity to quantities that are unrelated to the target quantity
- **Error:** difference between the sensor output m and the true value v

$$error = m - v$$

- **Accuracy:** degree of conformity between the sensor's measurement and the true value

$$accurance = 1 - \frac{|error|}{v}$$

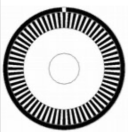
- **Precision:** reproducibility of the sensor results

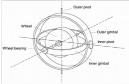
Sensor errors - challenges

- **Systematic errors:** caused by factors that can in theory be modeled; they are deterministic, e.g. calibration errors
- **Random errors:** cannot be predicted with sophisticated models; they are stochastic, e.g. spurious range-finding errors
- **Error analysis:** dperformed via a probabilistic analysis
  - Common assumption: symmetric, unimodal (and often Gaussian) distributions; convenient, but often a coarse simplification
  - Error propagation characterized by the *error propagation law*

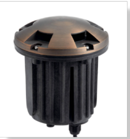
Ecosystem of sensors

- Encoders
- Heading sensors
- Gyroscope
- Accelerometers and IMUs
- Beacons
- Active ranging
- **Cameras**







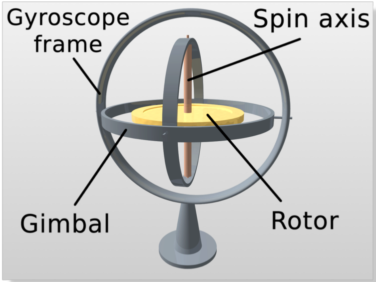

Encoders

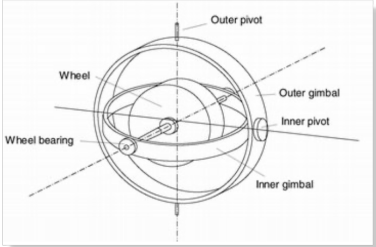
- **Encoder:** an electro-mechanical device that converts motion into a sequence of digital pulses, which can be converted to **relative** or **absolute** position measurements
    - proprioceptive sensor
    - can be used for robot localization
- **Fundamental principle of optical encoders:** use a light shining onto a photodiode through slits in a metal or glass disc

Wheel encoder
Credit: Pololu

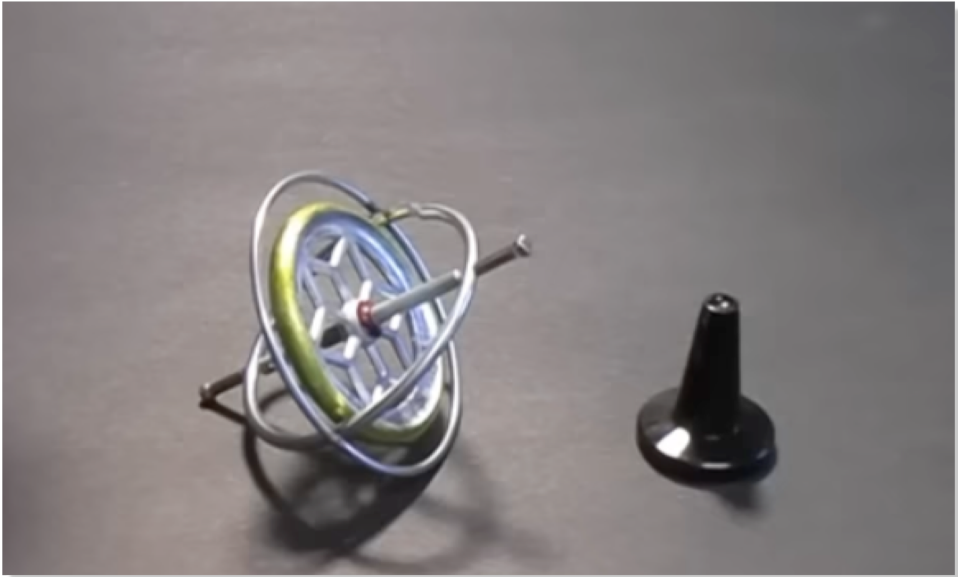
Heading sensors

- Used to determine robot's orientation, it can be:
  - Proprioceptive, e.g., **gyroscope** (heading sensor that preserves its orientation in relation to a fixed reference frame)
  - Exteroceptive, e.g., **compass** (shows direction relative to the geographic cardinal directions)
- Fusing measurements with velocity information, one can obtain a position estimate (via integration) $\rightarrow$ dead reckoning
- **Fundamental principle of mechanical gyroscopes:** angular momentum associated with spinning wheel keeps the axis of rotation inertially stable





Example Gyroscope



Source: https://youtu.be/cquvA_IpEsA?si=qTr_RIEppAkSyqc_, local video: Play Video

Accelerometer and IMU

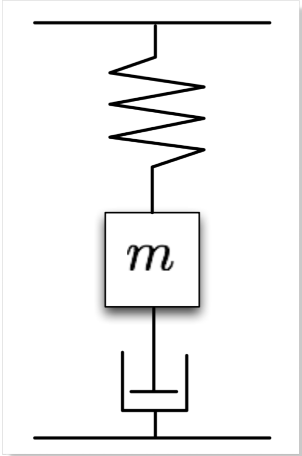
- **Accelerometer:** device that measures all external forces acting upon it
- Mechanical accelerometer: essentially, a spring-mass-damper system

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

with m mass of proof mass, c damping coefficient, k spring constant; in steady state

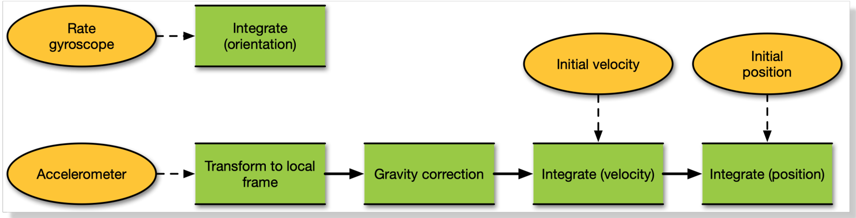
$$a_{applied} = \frac{kx}{m}$$

- Modern accelerometers use MEMS (cantilevered beam + proof mass); deflection measured via capacitive or piezoelectric effects



Inertial Measurement Unit (IMU)

- **Definition:** 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
- Drift is a fundamental problem: to cancel drift, periodic references to external measurements are required



Beacons

- **Definition:** signaling devices with precisely known positions
- Early examples: stars, lighthouses
- Modern examples: GPS, motion capture systems

Active ranging

- Provide direct measurements of distance to objects in vicinity
- Key elements for both localization and environment reconstruction
- Main types:
  - Time-of-flight active ranging sensors (e.g., ultrasonic and laser rangefinder)
  - Geometric active ranging sensors (optical triangulation and structured light)

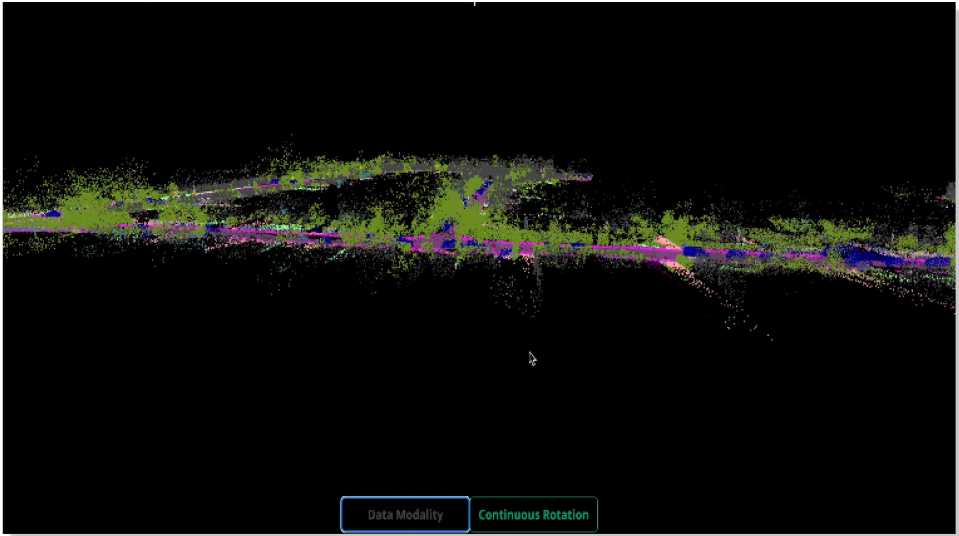
Time-of-flight active ranging

- **Fundamental principle:** time-of-flight ranging makes use of the propagation of the speed of sound or of an electromagnetic wave
- Travel distance is given by

$$d = ct$$

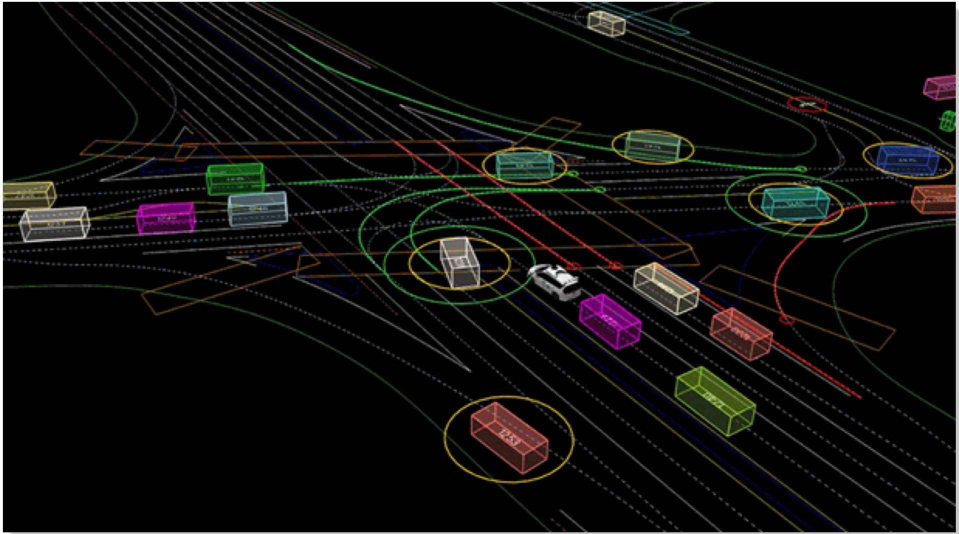
  where d is the distance traveled, c is the speed of the wave propagation, and t is the time of flight
- Propagation speeds:
  - Sound: 0.3 m/ms
  - Light: 0.3 m/ns
- Performance depends on several factors, e.g. uncertainties in determining the exact time of arrival and interaction with the target

Example Lidar data from Kitti 360 dataset



Source: https://www.thinkautonomous.ai/blog/lidar-datasets/

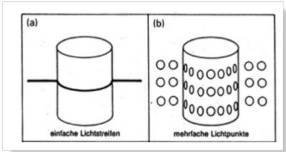
Example Lidar data from Waymo dataset



Source: `https://www.thinkautonomous.ai/blog/lidar-datasets/`

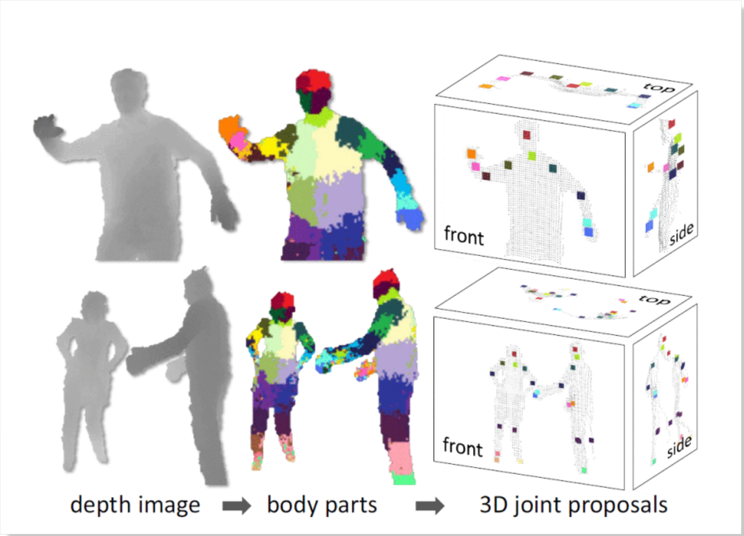
Geometric active ranging

- **Fundamental principle:** use geometric properties in the measurements to establish distance readings
- The sensor projects a known light pattern (e.g., point, line, or texture); the reflection is captured by a receiver and, together with known geometric values, range is estimated via triangulation
- Examples:
  - Optical triangulation (1D sensor)
  - Structured light (2D and 3D sensor)



Credit: Matt Fisher

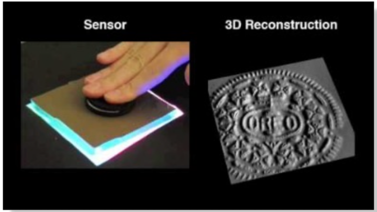
Real-Time Human Pose Recognition in Parts from Single Depth Images



Source: https://ieeexplore.ieee.org/document/5995316

Other sensors

- Classical, e.g. **Radar** (possibly using Doppler effect to produce velocity data, or **Tactile** sensors
- Emerging: **Artificial skin**, **Neuromorphic** cameras

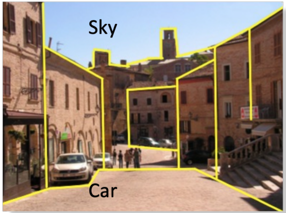


Computer Vision

- Aim:
  - Learn about cameras and camera models
- Readings:
  - Siegwart, Nourbakhsh, Scaramuzza. Introduction to Autonomous Mobile Robots. Section 4.2.3
  - D. A. Forsyth and J. Ponce [FP]. Computer Vision: A Modern Approach (2nd Edition). Prentice Hall, 2011. Chapter 1.
  - R. Hartley and A. Zisserman [HZ]. Multiple View Geometry in Computer Vision. Academic Press, 2002. Chapter 6.1.

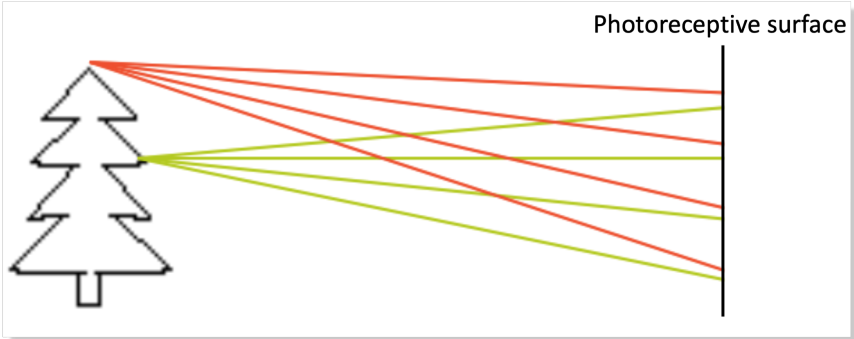
Vision

- Vision: ability to interpret the surrounding environment using light in the visible spectrum reflected by objects in the environment
- Human eye: provides enormous amount of information, millions of bits per second
- Cameras (e.g., CCD, CMOS): capture light → convert to digital image → process to get relevant information (from geometric to semantic)



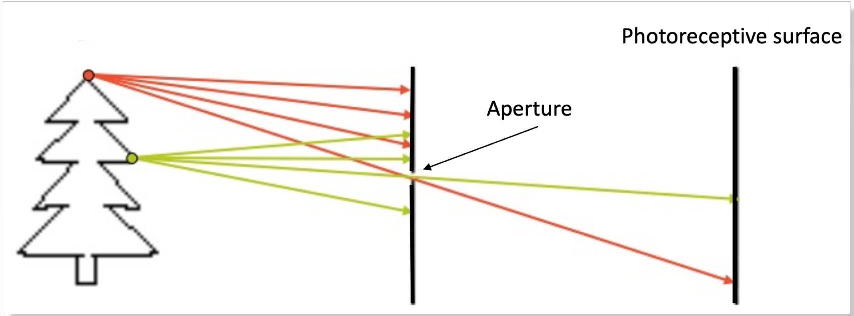
Capture an image of the world

- Light is reflected by the object and scattered in all directions
- If we simply add a photoreceptive surface, the captured image will be extremely blurred



Pinhole camera

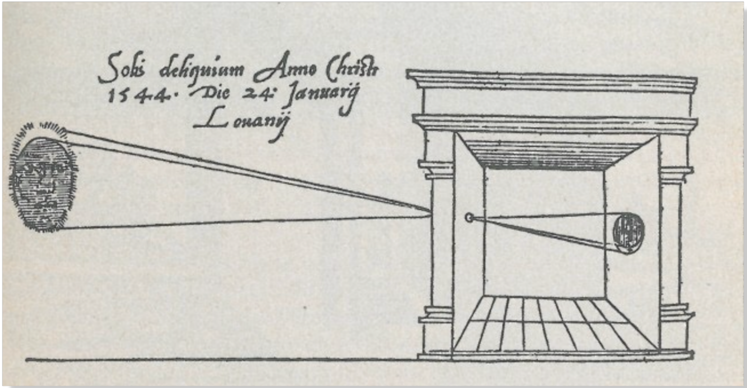
- **Idea:** add a barrier to block off most of the rays



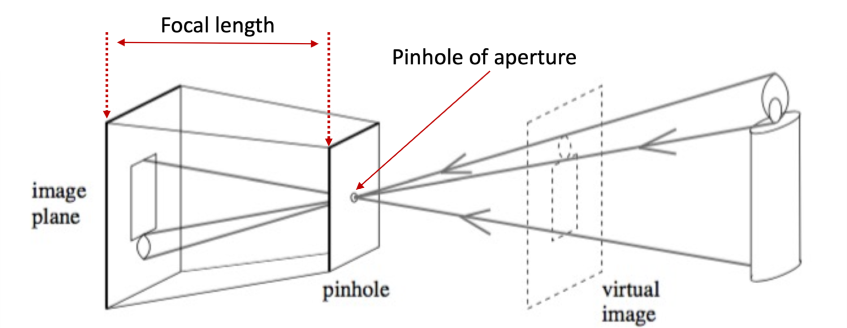
- **Pinhole camera:** a camera *without a lens* but with a tiny aperture, a *pinhole*

History

- Very old idea (several thousands of years BC)
- First clear description from Leonardo Da Vinci (1502)
- Oldest known published drawing of a camera obscura by Gemma Frisius (1544)



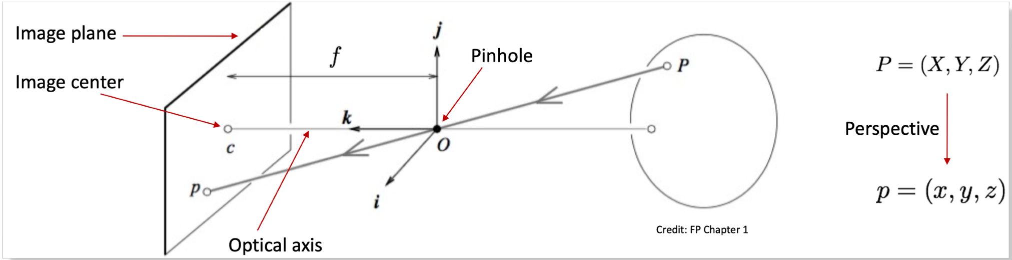
Pinhole camera



Credit: FP Chapter 1

- Perspective projection creates inverted images
- Sometimes it is convenient to consider a *virtual image* associated with a plane lying in front of the pinhole
- Virtual image not inverted but otherwise equivalent to the actual one

Pinhole perspective

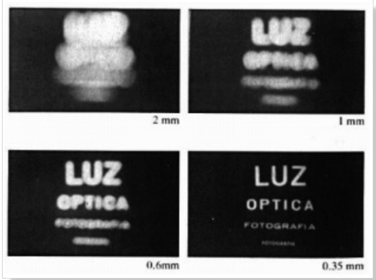


- Since P, O and p are collinear: $\bar{O}p = \lambda \bar{O}P$ for some $\lambda \in R$
- Also, $z = f$, hence

$$\begin{cases} x = \lambda X \\ y = \lambda Y \\ z = \lambda Z \end{cases} \Leftrightarrow \lambda = \frac{x}{X} = \frac{y}{Y} = \frac{z}{Z} \quad \Rightarrow \quad \begin{cases} x = f\frac{X}{Z} \\ y = f\frac{Y}{Z} \end{cases}$$

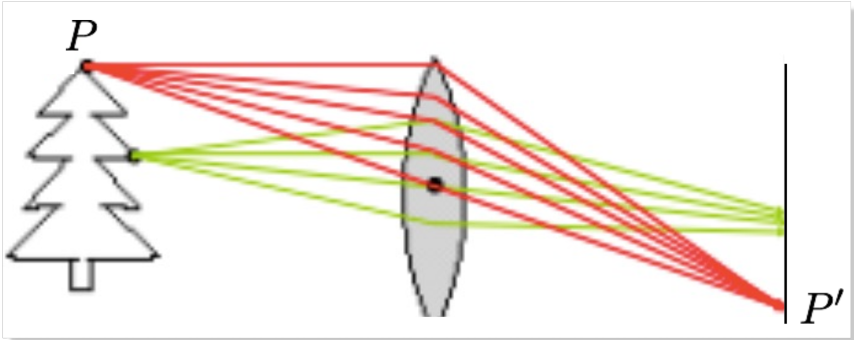
Issues with pinhole camera

- Larger aperture → greater number of light rays that pass through the aperture → blur
- Smaller aperture → fewer number of light rays that pass through the aperture → darkness (+ diffraction)
- **Solution:** add a lens to replace the aperture!

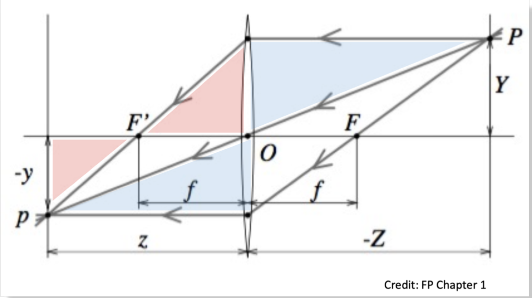


Lenses

- Lens: an optical element that focuses light by means of refraction



Thin lens model



- Similar triangles

$$\frac{y}{Y} = \frac{z}{Z} \qquad \textit{Blue triangles}$$

$$\frac{y}{Y} = \frac{z-f}{f} = \frac{z}{f} - 1 \qquad \textit{Red triangles}$$

Key properties (follows from Snell's law) :

- Rays passing through O are not refracted
- Rays parallel to the optical axis are focused on the *focal point* F'
- All rays passing through P are focused by the thin lens on the point p

$$\Rightarrow \frac{1}{z} + \frac{1}{Z} = \frac{1}{f} \qquad \textit{Thin lens equation}$$

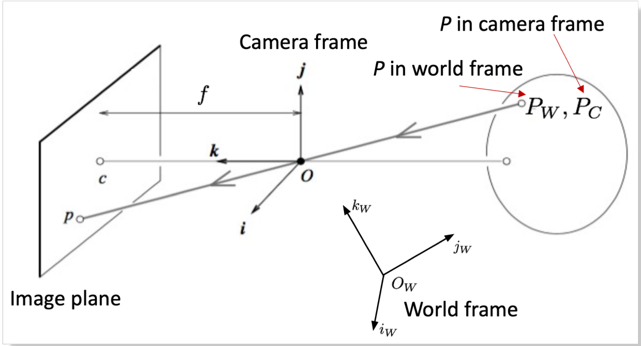
Thin lens model (2)

Key insights:

- **The equations relating the positions of P and p are exactly the same as under pinhole perspective if one considers z as focal length** (as opposed to f), since P and p lie on a ray passing through the center of the lens
- Points located at a distance $-Z$ from O will be in sharp focus only when the image plane is located at a distance z from O on the other side of the lens that satisfies the thin lens equation
- In practice, objects within some range of distances (called depth of field or depth of focus) will be in acceptable focus
- Letting $Z \to \infty$ shows that f is the distance between the center of the lens and the plane where distant objects focus
- In reality, lenses suffer from a number of aberrations

Perspective projection

- **Goal:** find how world points map in the camera image
- Assumption: pinhole camera model (all results also hold under thin lens model, assuming camera is focused at ∞)

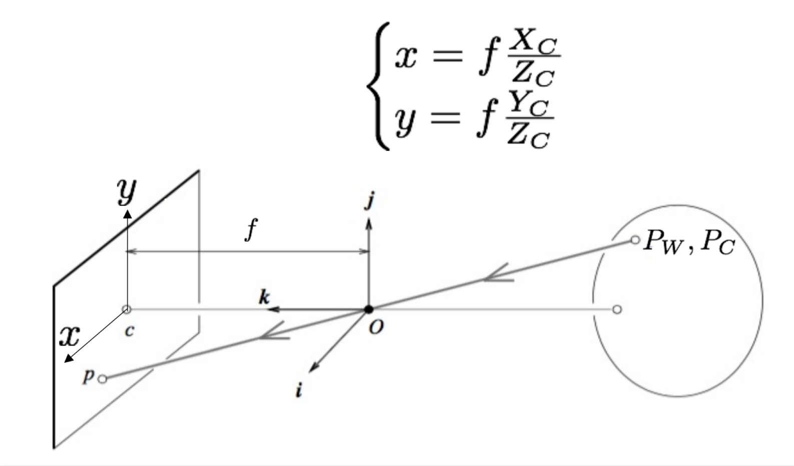


Roadmap:

- Map P_c into p (image plane)
- Map p into (u,v) (pixel coordinates)
- Transform P_w into P_c

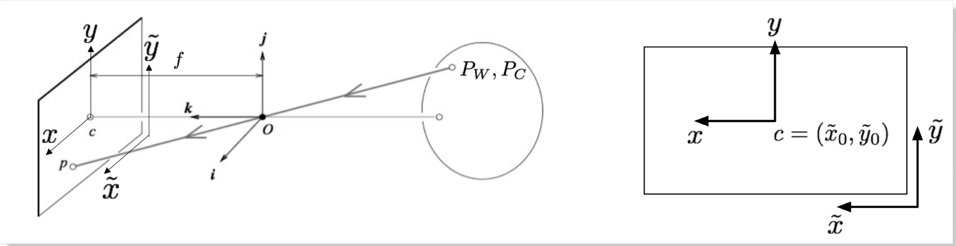
First step

- **Task:** Map $P_c = (X_C, Y_C, Z_C)$ into $p = (x, y)$ (image plane)
- From before



Second step (a)

- Actual origin of the camera coordinate system is usually at a corner (e.g., top left, bottom left)



Second step (b)

- Task: convert from image coordinates $(\tilde{x}, \tilde{y})$ to pixel coordinates (u, v)
- Let k_x and k_y be the number of pixels per unit distance in image coordinates in the x and y directions, respectively

$$u = k_x \tilde{x} = \overbrace{k_x f}^{\alpha} \frac{X_C}{Z_C} + \overbrace{k_x \tilde{x}_0}^{u_0}$$

$$v = k_y \tilde{y} = \underbrace{k_y f}_{\beta} \frac{Y_C}{Z_C} + \underbrace{k_y \tilde{y}_0}_{v_0}$$

$$\Rightarrow \quad u = \alpha \frac{X_C}{Z_C} + u_0, \qquad v = \beta \frac{Y_C}{Z_C} + v_0$$

Nonlinear transformation

Homogeneous coordinates

- Goal: represent the transformation as a linear mapping
- Key idea: introduce homogeneous coordinates

Inhomogenous -> homogeneous

$$\begin{pmatrix} x \\ y \end{pmatrix} \Rightarrow \lambda \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \qquad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \Rightarrow \lambda \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

Homogenous -> inhomogeneous

$$\begin{pmatrix} x \\ y \\ w \end{pmatrix} \Rightarrow \begin{pmatrix} x/w \\ y/w \end{pmatrix} \qquad \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} \Rightarrow \begin{pmatrix} x/w \\ y/w \\ z/w \end{pmatrix}$$

Perspective projection in homogeneous coordinates

- Projection can be equivalently written in homogeneous coordinates

$$\underbrace{\begin{bmatrix}\alpha & 0 & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1\end{bmatrix}}_{K}\begin{matrix}\\\end{matrix}\left[\begin{matrix}\alpha & 0 & u_0 & 0\\ 0 & \beta & v_0 & 0\\ 0 & 0 & 1 & 0\end{matrix}\right]\begin{pmatrix}X_c\\ Y_c\\ Z_c\\ 1\end{pmatrix} = \begin{pmatrix}\alpha X_c + u_0 Z_c\\ \beta Y_c + v_0 Z_c\\ Z_c\end{pmatrix}$$

Camera matrix/
Matrix of intrinsic parameters

P_c in homogeneous coordinates

Homogeneous pixel coordinates

- In homogeneous coordinates, the mapping is **linear**:

$$p^h = [K \quad 0_{3\times 1}] P_C^h$$

Point p in homogeneous pixel coordinates

Point P_c in homogeneous camera coordinates

Skewness

- In some (rare) cases

Skew parameter

$$K = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

- When is $\gamma \neq 0$?
  - x- and y-axis of the camera are not perpendicular (unlikely)
  - For example, as a result of taking an image of an image
- Five parameters in total!

Acknowledgements

Acknowledgement

This slide deck is based on material from the Stanford ASL and ETH Zürich

References

J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *CVPR 2011*, 2011, pp. 1297–1304.