Introduction to Mobile Robotics

Lecture 6: Perception

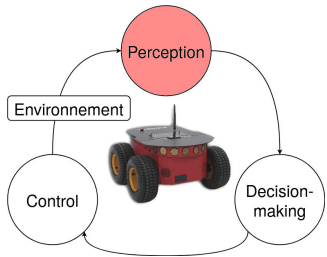
Zhi Yan

ENSTA, Institut Polytechnique de Paris

The Foundation of Autonomy

Definition:

Perception is the process by which a robot uses its sensors to acquire information about its environment and its own state, in order to make intelligent decisions.



The "sense" in sense-think-act

- Without accurate sensing, a robot is blind and unaware.
- The quality of its perception directly limits the quality of its decision-making and action.

ENSTA IP PARIS

The Challenge of Perception

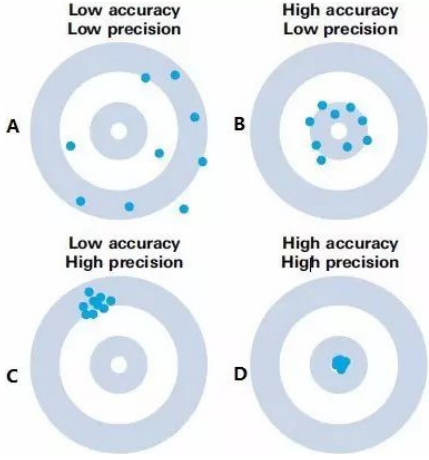
Why is perception hard?

- **Sensor noise:** Real-world measurements are never perfect.
- **Ambiguity:** Different world states can produce identical sensor readings.
- **Uncertainty:** A sensor provides only a partial, localized view of the world.
- **Unstructured data:** A point cloud is just a list of coordinates, an image is just a grid of pixels.

The goal of perception algorithms is to filter noise and reduce uncertainty to extract actionable information.

ENSTA IP PARIS

Sensors: Accuracy vs. Precision



Types of Robot Sensors

We can broadly classify sensors into two categories:

Proprioceptive sensors

- Measure the robot's **internal state**.
- Answer the question: "What is my own body doing?"
- Examples:
  - Wheel encoders
  - Inertial Measurement Unit (IMU)

Exteroceptive sensors

- Measure the **external environment**.
- Answer the question: "What is happening around me?"
- Examples:
  - Sonar / Ultrasonic
  - LiDAR (Light Detection and Ranging)
  - Cameras (Mono, Stereo, RGB-D)

A robust robot combines both types to get a complete picture.

ENSTA IP PARIS

Wheel Encoders: Measuring Motion

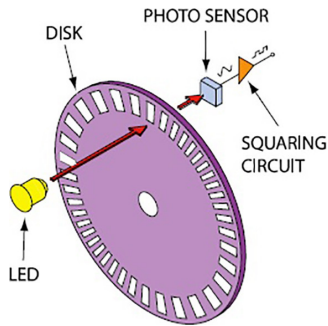
Wheel encoders measure the rotation of the robot's wheels.

$\Longrightarrow$ Foundation of **odometry** (estimation of motion based on actuator data).

How they work:

- A slotted disk is attached to the wheel axle.
- An LED shines light through the slots.
- A photodiode on the other side detects the light pulses.
- By counting pulses ("ticks"), we can precisely measure the angle of rotation.

$$\Delta\theta = (\text{ticks}) \times \frac{2\pi}{\text{ticks_per_revolution}}$$



ENSTA IP PARIS

Wheel Encoders: Measuring Motion

Limitation: Slippage

- Encoders measure wheel rotation, not the robot's actual movement.
- If a wheel slips (on a slick surface) or skids, the odometry will be inaccurate and errors will accumulate over time.
- This is known as **drift**.



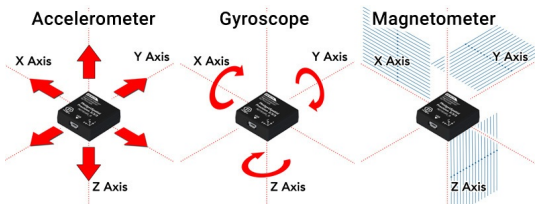
ENSTA
IP PARIS

IMU: Measuring Orientation

IMU (Inertial Measurement Unit) typically combines multiple sensors to measure a robot's orientation and angular velocity.

Core components:

- **Accelerometer:** Measures linear acceleration (including gravity).
- **Gyroscope:** Measures angular velocity ($\omega_x, \omega_y, \omega_z$).
- **Magnetometer:** Measures the local magnetic field (like a compass).



Let's take a closer look at each one.

IMU: The Accelerometer

Principle: Mass-on-a-Spring

Infer acceleration via Hooke's Law ($F = kx = ma$).

Measuring tilt:

- At rest, the only acceleration an IMU measures is gravity ($g \approx 9.8 m/s^2$) pointing down.
- By measuring the components of this gravity vector along the robot's X, Y, and Z axes, it can determine tilt (roll and pitch angles).
- Roll (ϕ) $= \text{atan2}(A_y, A_z)$
- Pitch (θ) $= \text{atan2}(-A_x, \sqrt{A_y^2 + A_z^2})$

Limitation:

- Cannot measure yaw (rotation around the gravity vector).
- Prone to noise from vibrations.

ENSTA IP PARIS

IMU: The Gyroscope

Principle: Coriolis[1] effect

- Modern MEMS gyroscopes use a vibrating structure.
- When the robot rotates, the Coriolis effect induces a secondary vibration perpendicular to the primary one.
- The magnitude of this secondary vibration is proportional to the angular velocity.

Tracking orientation:
Integrate the angular velocity over time: $\text{angle}_{\text{new}} = \text{angle}_{\text{old}} + \omega \cdot \Delta t$

Limitation:

- Every measurement of ω has a small error.
- When integrated, grows into a large orientation error.

[1] (Gustave Coriolis, X 1808)

ENSTA IP PARIS

IMU: The Magnetometer & Fusion

Magnetometer:

- Measures the local magnetic field (like a compass).
- Provides an absolute heading (yaw) relative to magnetic north.
- Limitation: Can be easily disturbed by nearby metal (steel beams in buildings) or electrical currents from the robot's own motors.

Fusion:

No single sensor is sufficient. We combine them:

- **Accelerometer:** Good for long-term tilt (roll, pitch), but noisy.
- **Gyroscope:** Good for short-term rotation (all axes), but drifts.
- **Magnetometer:** Good for long-term heading (yaw), but easily disturbed.

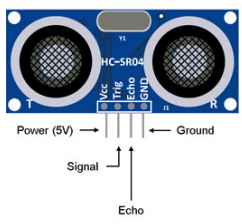
Data from these sensors is almost always **fused** *(e.g., using a Kalman Filter) to provide a more stable and accurate orientation estimate.*

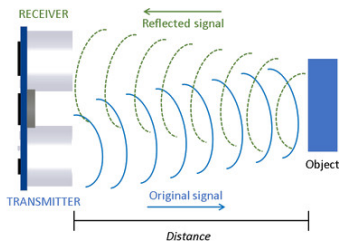
ENSTA IP PARIS

Sonar / Ultrasonic

Principle: Time-of-Flight (of sound)

1. An emitter releases a high-frequency pulse of sound (often called a "ping").
2. The sound travels, hits an object, and reflects back.
3. A "microphone" measures the exact time (Δt) it took for the echo to return.
4. Distance is calculated: $d = \frac{v_{sound} \cdot \Delta t}{2}$.





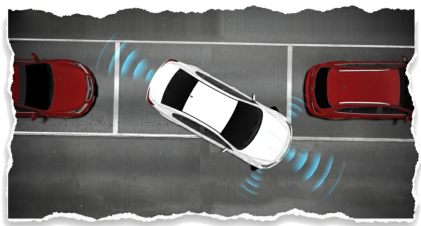
Sonar / Ultrasonic

Pros:

- Very inexpensive
- Low power consumption
- Can detect transparent objects (like glass) that are invisible to LiDAR

Cons:

- Wide detection cone (poor angular resolution)
- Prone to specular reflections
- Slow update rate



Primary use: Basic, low-cost obstacle detection (e.g., car parking sensors).

ENSTA IP PARIS

LiDAR

Principle: Time-of-Flight

1. A laser emits a short, intense pulse of light.
2. The pulse travels, hits an object, and reflects back.
3. A highly sensitive detector measures Δt it took for the pulse to return.
4. Distance is calculated: $d = \frac{c \cdot \Delta t}{2}$ (where c is the speed of light: $3 \times 10^8 m/s$).

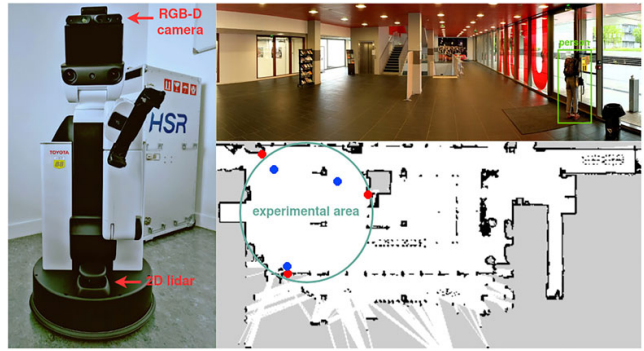
LiDAR can build a map of the environment by measuring distances!

ENSTA
IP PARIS

LiDAR

2D LiDAR:

- A single laser beam is spun around by a rotating mirror.
- Produces a 2D planar "slice" of the environment.
- Common in indoor robots.
- In ROS 2, publishes `sensor_msgs/msg/LaserScan`.



The LaserScan Message

`sensor_msgs/msg/LaserScan`: A fundamental message type in ROS 2 for 2D LiDAR data.

Key fields:

```
std_msgs/Header header
float32 angle_min          # start angle of the scan [rad]
float32 angle_max          # end angle of the scan [rad]
float32 angle_increment    # angular distance between measurements [rad]
float32 time_increment     # time between measurements [seconds]
float32 scan_time          # time between scans [seconds]
float32 range_min          # minimum range value [m]
float32 range_max          # maximum range value [m]
float32[] ranges           # range data [m] (the core data!)
float32[] intensities      # intensity data
```

The “ranges” array contains the distance measurements. For a 360-degree scan with 1-degree resolution, this array will have 360 elements. The first element corresponds to “angle_min”, the next to “angle_min + angle_increment”, and so on.

ENSTA IP PARIS

Visualizing LiDAR Data

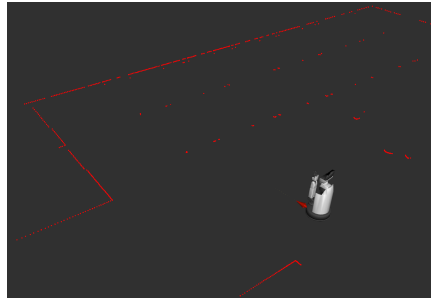
How do we make sense of a "LaserScan"?

The "ranges" array is just a list of numbers. To understand it, we must visualize it in a coordinate frame. Each measurement "ranges[i]" is a point in polar coordinates (r, θ) where:

- $r = \text{ranges}[i]$
- $\theta = \text{angle_min} + i \times \text{angle_increment}$

These are then converted to Cartesian coordinates (x, y) for plotting:

- $x = r \cdot \cos(\theta)$
- $y = r \cdot \sin(\theta)$

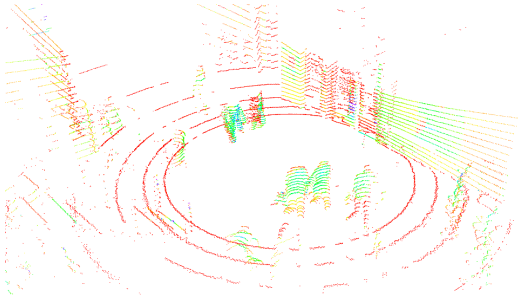


In our practical work, you will use this data to detect obstacles and navigate a maze.

ENSTA IP PARIS

3D LiDAR

- Uses multiple lasers (e.g., 16, 32, 64) stacked vertically, all rotating together.
- Produces a full 3D point cloud of the environment.
- The standard for autonomous vehicles and advanced robotics.
- In ROS 2, publishes `sensor_msgs/msg/PointCloud2`.



Pros: Very accurate, long range, works in any lighting condition.
Cons: Expensive, can fail on transparent or black surfaces, struggles in bad weather (rain, fog, etc.).

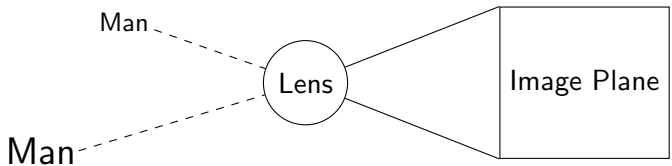
ENSTA IP PARIS

Cameras

Cameras provide dense, rich information about the world (color, texture, shape) but do not directly measure distance.

Monocular camera:

- Provides a 2D image.
- In ROS 2, publishes `sensor_msgs/msg/Image`.
- **Challenge:** Cannot recover depth from a single image.
- **Technique:** Deep learning can be used to *estimate* depth, but still not a direct measurement.



A large object far away can look the same as a small object up close.

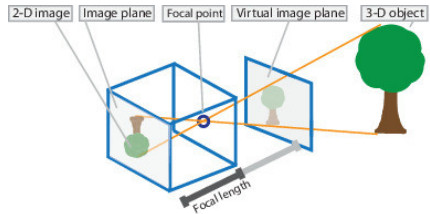
The Pinhole Camera Model

A point (X, Y, Z) in the 3D world maps to a pixel (u, v) on the 2D image plane.

Projection equations (simplified):

$$u = f_x \frac{X}{Z} + c_x$$

$$v = f_y \frac{Y}{Z} + c_y$$



Where:

- (X, Y, Z) is the 3D point's coordinate.
- (f_x, f_y) is the focal length in pixels.
- (c_x, c_y) is the principal point (image center).

Notice that the depth, Z, is in the denominator. $\Longrightarrow$ *The mathematical reason why depth information is lost in a single 2D image.*

ENSTA IP PARIS

Stereo Cameras

Uses two cameras separated by a known distance (the baseline), mimicking human binocular vision:

1. Find the same feature point in both the left and right images.
2. The difference in the horizontal position of the point in the two images is called **disparity**.
3. Using triangulation, depth can be calculated: $Depth \propto \frac{Baseline}{Disparity}$.
4. The larger the disparity, the closer the object.

Pros: Passive (doesn't emit light), provides dense depth.
Cons: Computationally intensive, struggles on textureless surfaces (e.g., a white wall), limited effective range.

ENSTA IP PARIS

RGB-D Cameras

An RGB-D camera provides two synchronized images: a regular color (RGB) image and a per-pixel depth (D) image.



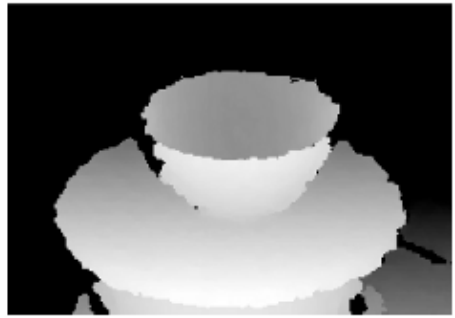


Image source: `https://rgbd-dataset.cs.washington.edu/`

ENSTA IP PARIS

RGB-D Cameras

Structured light

- Projects a known pattern of infrared light onto the scene.
- An infrared camera observes how the pattern is distorted by the objects.
- Depth is calculated based on the distortion.
- *Example: Original Microsoft Kinect.*

Time-of-Flight

- An array of infrared emitters sends out modulated light.
- The sensor measures the phase shift of the returning light for each pixel.
- This phase shift corresponds to distance.
- *Example: Azure Kinect, modern phone LiDARs.*

ENSTA
IP PARIS

RGB-D Cameras

Structured light

Pros:

- High accuracy at close range
- Potentially lower cost

Cons:

- Easily washed out by sunlight
- Baseline between projector/camera creates minimum range
- Can be computationally heavier

Time-of-Flight

Pros:

- Less computationally expensive
- More compact design
- Works better in bright ambient light

Cons:

- Can fail on reflective surfaces
- Can suffer from multi-path interference
- Limited range

Time-of-Flight is becoming the dominant technology for modern RGB-D sensors.

ENSTA IP PARIS

Processing Point Clouds

A raw 3D point cloud from a LiDAR or RGB-D camera can contain millions of points. It is often too noisy and dense to be used directly.

Common pre-processing steps:

- **Downsampling:** Reduce the number of points to speed up computation (e.g., using a Voxel Grid filter).
- **Outlier removal:** Remove stray points caused by measurement errors (e.g., using a Statistical Outlier Removal filter).
- **Segmentation:** Group points into meaningful clusters or surfaces (e.g., using RANSAC to find the ground plane or walls).
- **Normal estimation:** Calculate the surface normal vector for each point, which is useful for object recognition and registration.

ENSTA
IP PARIS

Processing Images

An image is a grid of pixels. For tasks like visual SLAM or object recognition, a robot must identify distinctive, re-identifiable points called **features**.

What makes a good feature?

A point in an image (like a corner) that is salient and can be reliably found again from different viewpoints and under different lighting conditions.

Feature detectors and descriptors:

- **Detector:** Finds the (u, v) locations of interesting keypoints. Examples: Harris Corner, SIFT, SURF, ORB.
- **Descriptor:** For each keypoint, it computes a vector of numbers that summarizes the appearance of the image patch around it. This vector is used for matching features between images.

ENSTA
IP PARIS

No Perfect Sensor

- Every sensor has strengths and weaknesses.
- The key to robust perception is to combine multiple, complementary sensors.
- This is called **sensor fusion**.

$\Longrightarrow$ The core idea of SLAM (next lecture)!

ENSTA
IP PARIS

Sensor Comparison Summary

Sensor	Measures	Pros	Cons	Main use
Encoders	wheel rotation	precise, high-freq	slips, accumulates error	odometry
IMU	orientation, ang. vel.	fast, provides attitude	drifts badly, noisy	orientation, odometry fusion
LiDAR	precise distances	accurate, all-lighting	expensive, specific surfaces	mapping, localization, avoidance
Camera	light intensity (color)	rich data, cheap	no direct depth, lighting	object recognition, V-SLAM
Stereo cam	disparity $\rightarrow$ depth	passive depth	textureless surfaces, cpu	depth perception, obstacles
RGB-D cam	per-pixel depth	direct depth, cheap	sunlight, range, reflections	indoor navigation, HRI

ENSTA
IP PARIS

Questions?

Next: Practical Work 6 - Maze Solving

ENSTA
IP PARIS