Coordinate frames on/for robots

What we will do today ..

• Assignment No. 1, Wednesday (9/18)
• Odometry ($u_i$ in Fig. 3 of [Caneda 2016])
• Wheel odometry
• LiDAR and LiDAR odometry
• Inertial measurement unit (IMU) and IMU odometry
• Cameras (RGB and RGB-D)
• On Wed (9/25), we will talk about loop closure ($c_i$) the course project.

Odometric constraint: $u_i(z_k)$

From the kinematics of differential drive robot (pp. 4-5), let:
$$x_{i+1} = [x', y', \theta']^T, x_i = [x, y, \theta]^T, u_i = [v_l, v_r]^T$$
the robot kinematic equation is of the form: $x_{i+1} = f(x_i, u_i)$. If we define $x_k = [x_k, x_{k+1}]^T$ and $z_k = u_0$, our SLAM optimization problem needs to be generalized to
$$X^* = \arg \min_{X} \sum_{k=0}^{m} g_k(x_k, x_{k+1})_i$$
where
$$g_k(x_k, x_{k+1}) = x_{i+1} - f(x_i, u_i).$$

In 3D, we need to change the above cost term into a matrix form (see the document 'g2o versus Toro: Format and Cost Functions'):
$$\min_{X} \sum_{(i,j) \in \mathcal{E}} \| R_x R_y (x_{i+1} - f(x_i, u)) \|_2^2 + \| u_j - R_i (t_i - x) \|_2^2$$

Odometric constraint: $u_i(z_k)$

From the kinematics of differential drive robot (pp. 4-5), let:
$$x_{i+1} = [x', y', \theta']^T, x_i = [x, y, \theta]^T, u_i = [v_l, v_r]^T$$
the robot kinematic equation is of the form: $x_{i+1} = f(x_i, u_i)$. If we define $x_k = [x_k, x_{k+1}]^T$ and $z_k = u_0$, our SLAM optimization problem needs to be generalized to
$$X^* = \arg \min_{X} \sum_{k=0}^{m} g_k(x_k, x_{k+1})_i$$
where
$$g_k(x_k, x_{k+1}) = x_{i+1} - f(x_i, u_i).$$

In 3D, we need to change the above cost term into a matrix form (see the document 'g2o versus Toro: Format and Cost Functions'):
$$\min_{X} \sum_{(i,j) \in \mathcal{E}} \| R_x R_y (x_{i+1} - f(x_i, u)) \|_2^2 + \| u_j - R_i (t_i - x) \|_2^2$$
LiDAR – Light Detection and Ranging

Function and Concept of 2D LiDAR

All LiDAR units operate using this basic set of steps:
1. Laser light is emitted from the unit (usually infrared).
2. Laser light hits an object and is scattered.
3. Some of the light makes it back to the emitter.
4. The emitter measures the distance (more on how later).
5. Emitter turns, and the process begins again.

3D LiDAR is built from multiple planes or a pan-tilt mirror system.

Types of LiDAR

How exactly the laser sensor measures the distance to an object depends on how accurate the data needs to be.

1. Time of Flight:
   \[ \text{distance} = \text{(speed of light)} \times \text{(time)} \]

2. Phase Difference (used by robots):
   modulate laser beam and measure phase shift (e.g. 250 KHz)

3. Angle of incident:
   infer distance from the angle of reflected light

LiDAR Selection Criteria

- Range
- Light sensitivity
- Angular resolution
- Field of View
- Refresh rate (a few times/s)
- Accuracy – noise in range measure
- Size
- Cost

HDL-64E

$100,000!

2D LiDAR Scan from a SICK sensor
LiDAR Scan from Velodyne HDL-64E

LiDAR Odometry [Zhang et al. 2014]

Computes the motion of the LiDAR between two consecutive scans.

1. Select feature points that are on either sharp edges or planar surface patches

\[ e = \frac{1}{|S|} \sum_{(i,j) \in S} \| X_i - X_j \| \]

2. Match edge points and planar points in consecutive scans.

3. Estimate LiDAR motion from matched points.

Step 3: Iterative Closest Point (ICP)

Find a rigid transform \( T = [R, t] \) to transform one set of points (e.g. blue) so that the sum of the distances between the corresponding points in the two sets is minimized.
What is in an IMU?

- Gyroscopes -> Angular Velocity (radians/s)
- Accelerometer -> Linear Acceleration (m/s² or g)
- Magnetometer -> Magnetic field strength (micro-Tesla or Gauss)

In addition, sometimes, barometric or GPS measurements are integrated in an IMU as well.

Gravity

When $m$ accelerates, it will experience a force, which will cause spring $K$ to move. The displacement of $k$ can be measured through sensor $x$, and the force $F$ can be sensed as a result. Given $F$ and $m$, the acceleration $\ddot{x}$ can be measured based on Newton's 2nd law.

$$\ddot{x} = \frac{F}{m}$$

Three (3) accelerometers along the x, y, and z axes provide three linear accelerations whose double-integration results in linear displacement measurements (odometry).

Newton's first law of motion states: a body in motion will remain in motion unless it is acted upon by an external force.
Gyroscope’s Principle of Operation

Imagine a particle traveling in space with a velocity vector \( \vec{V} \). An observer sitting on the x-axis in Fig. (a) is watching this particle. If the coordinate system along with the observer starts to rotate around the z-axis with an angular velocity \( \vec{\Omega} \), the observer thinks that the particle is changing its trajectory toward the x-axis at an acceleration

\[
2 \times \vec{V} \times \vec{\Omega} \quad \text{or} \quad \text{Coriolis force} \quad F = m \times 2 \times \vec{V} \times \vec{\Omega}.
\]

(Navid Yazdi et al., "Micromachined inertial sensors", Proc. IEEE, 1998.)

Gyroscope: tuning fork design

Tuning fork design uses two tines that resonate \( \vec{V} \). Coriolis force is detected either as differential bending of the tines or as a torsional vibration of the tuning-fork stem, along a direction that is orthogonal to the main vibration. From the Coriolis force, angular velocity \( \vec{\Omega} \) can be calculated.

Gyroscope or Gyro

Coriolis force is detected either as differential bending of the tuning fork tines or as a torsional vibration of the tuning-fork stem, along a direction that is orthogonal to the main vibration.

Gyroscope or Gyro

Three gyroscopes along three principal axes provide three angular velocities whose integration results in angular change (odometry). Modern gyros are built with MEMS (right).

Gyroscope applications

Gyros are responsible for UAV control, ESC (electronic stability control) for cars to prevent rollovers, Wii by Nintendo, video games on smart phones, etc.

Summary

- IMU uses gyroscopes to measure angular velocities and accelerometers to measure accelerations, about the three axes.
- Integration of angular velocity over a short time interval provides angular change.
- Double integration of acceleration over a short time interval provides linear change.
- IMU is inexpensive but requires calibration and is subject to ambient noise in environments.
Cameras and Images

- Why camera?
- Images and Pixels
- Image formation
- Coordinate frames: image, camera, and world (map)
- Camera calibration: intrinsics and extrinsics

Why camera?

- Vision provides rich information about the world in which a robot operates (humans derive 80-90% of information from vision).
- Types of information computed with vision in robotics:
  - Geometry of a scene (SfM, single-image depth, surface normal)
  - Robot motion (visual odometry)
  - Object detection (pedestrians, cars, doors, windows, etc.)
  - Object classification (scene semantics)

Images and pixels

- A camera captures either color or grayscale images
- Each grayscale image is a matrix of N columns and M rows (e.g., 640x480)
- A color image consists of three separate images, one for each color component or channel (e.g., RGB)
- A color image can be captured with one array of lighting sensing elements covered with a Bayer filter mosaic or three separate arrays.
- Each element of an image array is called a picture element or pixel, indexed by its column and row number.
- The intensity (grayscale or a color channel) value of a pixel is encoded typically in 8-bits (0-255).
- Therefore, a grayscale image is just an array of 8-bit numbers, which a computer vision algorithm processes to extract info.

Image formation: pinhole model

- A camera captures either color or grayscale images
- Each grayscale image is a matrix of N columns and M rows (e.g., 640x480)
- A color image consists of three separate images, one for each color component or channel (e.g., RGB)
- A color image can be captured with one array of lighting sensing elements covered with a Bayer filter mosaic or three separate arrays.
- Each element of an image array is called a picture element or pixel, indexed by its column and row number.
- The intensity (grayscale or a color channel) value of a pixel is encoded typically in 8-bits (0-255).
- Therefore, a grayscale image is just an array of 8-bit numbers, which a computer vision algorithm processes to extract info.
Pinhole camera

Image coordinate vs camera cameraframe

World coordinate frame

Camera parameters

Coordinate frames on/for robots