Introduction to Intelligent Vehicles
[8. Sensing and Perception]

Chung-Wei Lin

cwlin@csie.ntu.edu.tw

CSIE Department

National Taiwan University

We would like to thank Prof. Jyo Deshmukh (USC) for contributing the lecture material

Layered View of Autonomous Vehicles

Sensing	Perception	Planning and Decision	Control	Actuation
Radar	Segmentation	Route Planning	Steering Control	Engine
Lidar	Object Detection	Behavior Planning	Torque Control	Brake
Camera	Object Tracking	Motion Planning	Emission Control	Wheel
GPS	Localization	Prediction	Energy Management	Light

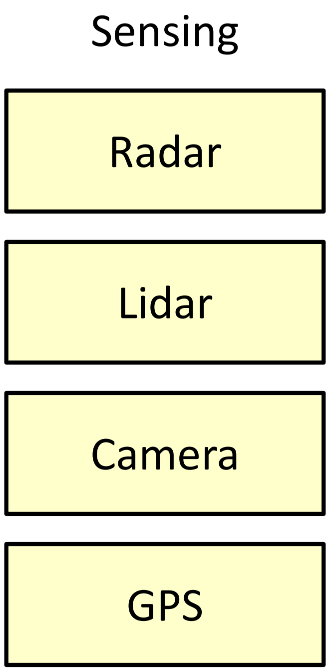
(Operating System)

Hardware Platform, e.g., Electronic Control Unit (ECU)

Sensing

- Detection of a physical presence [Wikipedia]
- Conversion of that data into a signal that can be read by an observer or an instrument [Wikipedia]

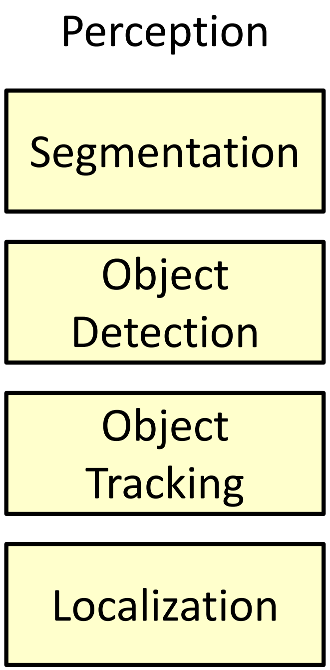
- Most software in sensing deals with sensor fusion and data preprocessing
  - Converting lidar data into point-cloud representation
  - Sampling video signals
  - Compressing data



Perception

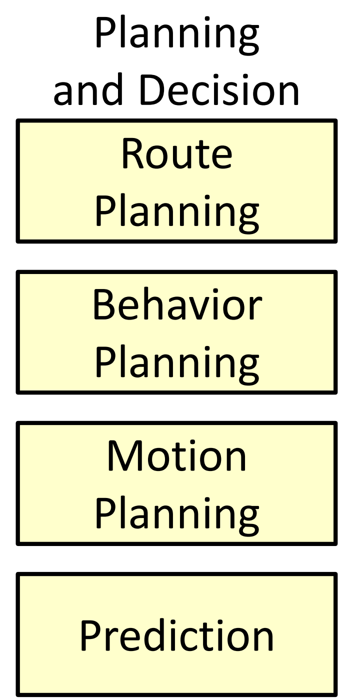
- Organization, identification, and interpretation of sensory information in order to represent and understand the presented information, or the environment [Wikipedia]

- Many things fall under the vague category of perception
  - Strongly connected to sensor fusion
  - Providing "meaning"



Planning and Decision

- The most software-intensive layer
- Several algorithms have been proposed in the robotics and automotive communities based on
  - Optimization
  - Search-based planning
  - Discrete decision-making (with state machines)
- Current trend is to investigate application of AI/control techniques such as
  - Reinforcement learning
  - Deep learning



Control

- Developed before autonomous vehicles
- Recent trends
  - "Drive-by-wire"
    - Replace mechanical and hydraulic components by electrical and electronic components
  - More efficient control with data
    - Models of the environment

Control

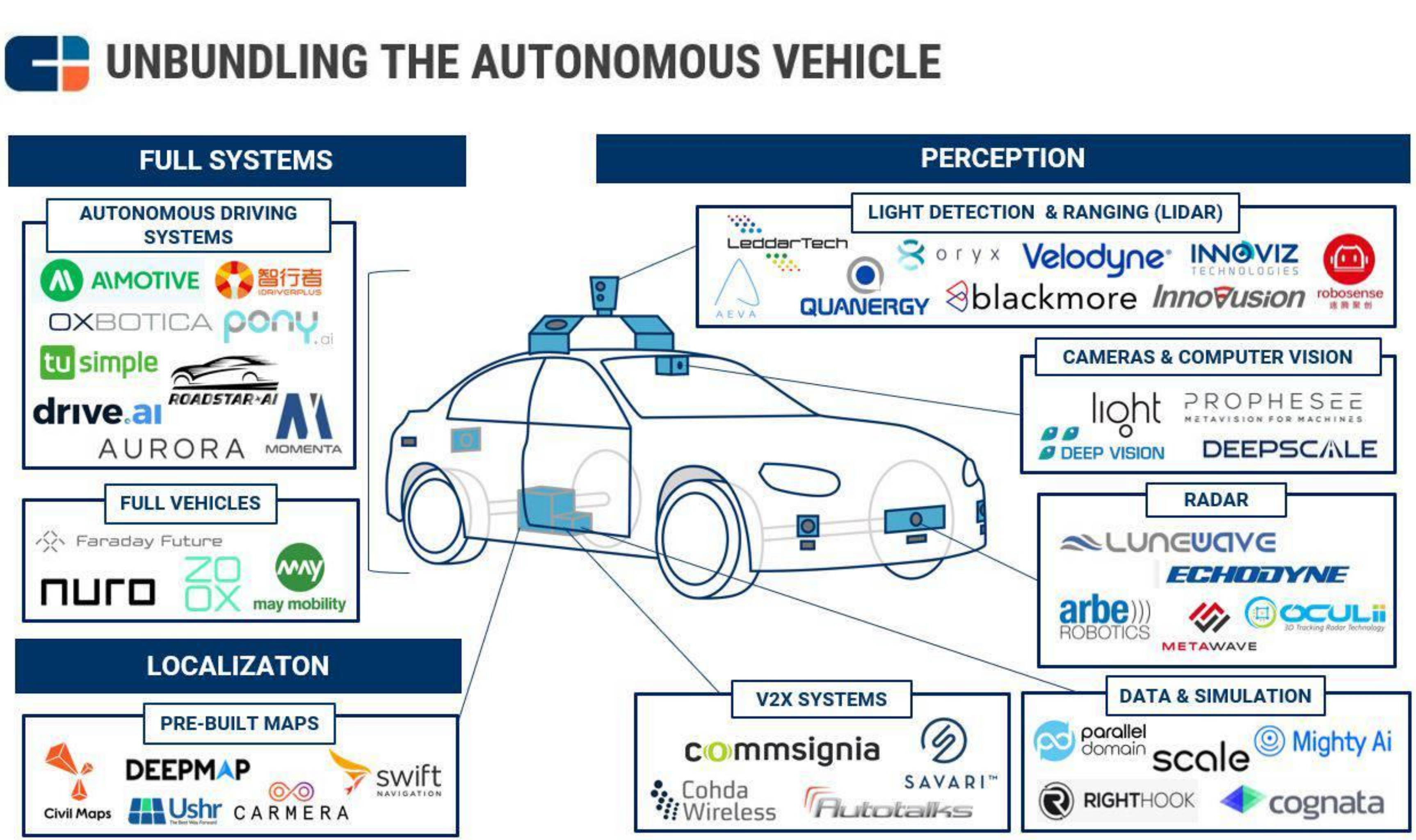
Steering Control

Torque Control

Emission Control

Energy Management

Battlefield



https://www.cbinsights.com/research/startups-drive-auto-industry-disruption/?fbclid=IwAR33R9kYhypQ2GkN9Dsd8cs2Zyu5K3SWLH0i9WTsm1pdjqZSZhfC1FQtztk

Outline

- ❑ **<u>Sensing</u>**
- ❑ Perception

Sensors

- Transducers that convert one physical property into another
  - In our context, a sensor converts a physical quantity into a numeric value
- Examples
  - Radar
  - Lidar
  - Ultrasonic
  - Camera
  - Global Positioning System (GPS)
  - Inertial Measurement Unit (IMU)
  - Others?

Sensor Selection

- Selection of sensors is a hard design problem, as sensors can have many factors that need tuning
  - Accuracy
    - Error between true value and its measurement
  - Resolution
    - Minimum difference between two measurements
  - Sensitivity
    - Smallest change in value that can be detected
  - Sensing range (perspective)
    - Which portion of the environment can the sensor measure (e.g., the field of vision for a camera, orientation for an ultrasonic module)
  - Data range
    - Minimum and maximum values that can be accurately detected
  - Frequency and responsiveness
  - Interface

Radar

- One major principle for "motion" is the Doppler effect
- Continuous-wave radar
  - Unmodulated continuous-wave radar
    - Sense velocity only
  - Modulated continuous-wave radar
    - Sense both of distance and velocity
    - Frequency-modulated continuous-wave radars are popular
- Radars may require additional signal processing to give precise answers when the environment is dusty, rainy, or foggy
- Forward-facing radars estimate relative position and velocity of the lead vehicle
- Cheap, low resolution, good in extreme weather (compared with lidar and camera)

Lidar

- Lidar stands for "light detection and ranging"
- Typical lidars use multi-beam light rays ("ray-casting")
  - A lidar casts a ray at an angle
  - The first obstacle reflects the ray
  - The lidar gets the distance from the first obstacle
- Lidar data consists of rotational angles and distances to obstacles
  - This can be represented in a point cloud form by mapping each obstacle point to 3D coordinates
- Expensive, higher resolution (compared with radar), extremely accurate depth information

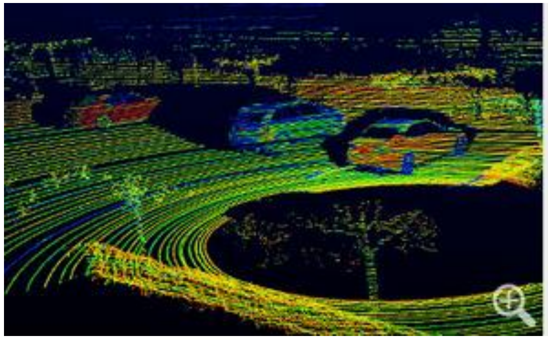
Lidar: Example 1

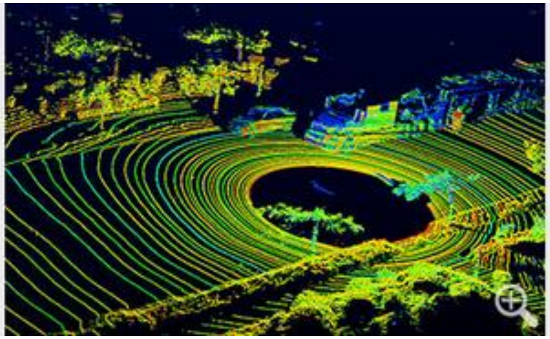
https://www.hokuyo-aut.jp/search/single.php?serial=167

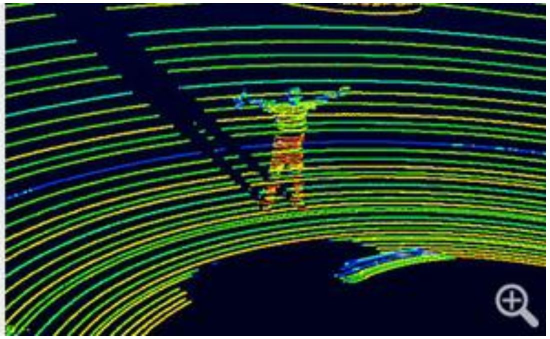
- Hokuyo UST-10LX
  - Light source: semiconductor laser diode
  - Scanning range: 0.02-10m, 270°
  - Measuring accuracy: ±40mm
  - Angular resolution: 0.25°
  - Scanning frequency: 40Hz
  - Communication: SCIP 2.2 and Ethernet 100 BASE-TX
  - Power source: 12V or 24VDC
  - Power consumption: 0.15A or less (on 24VDC)
  - Weight: 130g
  - Size: 50 x 50 x 70 mm^3
  - Avoid direct sunlight as it may cause sensor malfunction

Lidar: Example 2

- Velodyne HDL-64E
  - 64 channels
  - 120m range
  - Up to ~2.2 million points per second
  - 360° horizontal Field Of View (FOV)
  - 26.9° vertical FOV
  - 0.08° angular resolution (azimuth)
  - ~0.4° vertical resolution
  - User selectable frame rate







https://velodynelidar.com/hdl-64e.html

Ultrasonic

- Ultrasound is sound waves with frequencies higher than the upper audible limit of human hearing [Wikipedia]
  - Ultrasonic devices are used to detect objects and measure distances

- Example
  - Ultrasonic Range Finder - XL-MaxSonar-EZ4
    - Range: 0 to 765cm
    - Resolution: 1cm

Camera

- Cheap, higher resolution (compared with radar and lidar), inaccurate depth information, not good in extreme weather

- Example: Mobileye
  - https://www.youtube.com/watch?v=dhEgD6ZFlQE

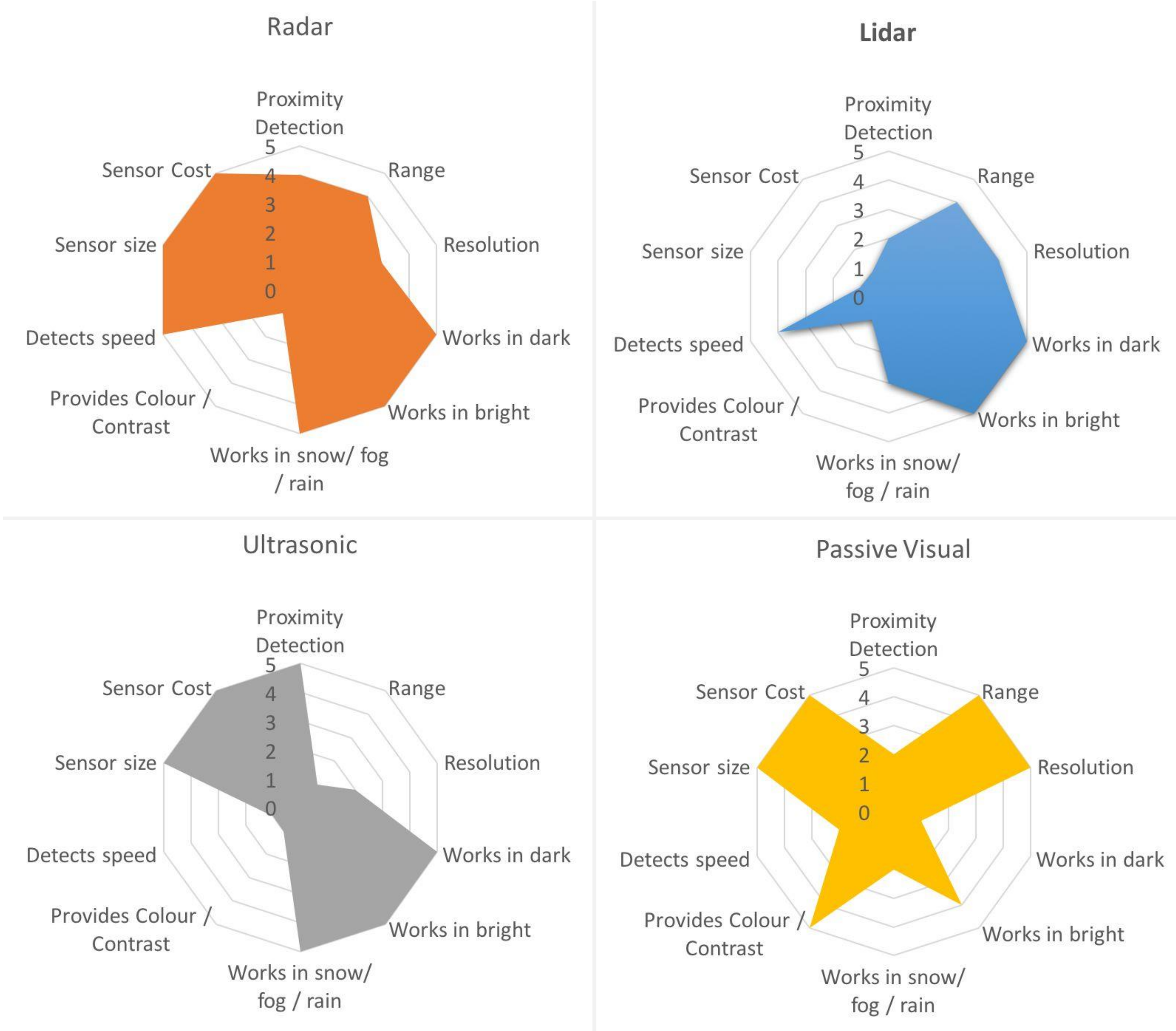
GPS

- GPS gives information about **current time**, latitude, longitude, and altitude
- Commercial GPS systems provide (GPS) coordinates of a vehicle within ±2m accuracy
  - This is not enough for autonomous driving
  - Some GPS receivers provide decimeter level accuracy
- Often GPS data is used as "observations" along with an IMU-based inertial navigation system (INS) to localize the vehicle

Inertial Measurement Unit (IMU)

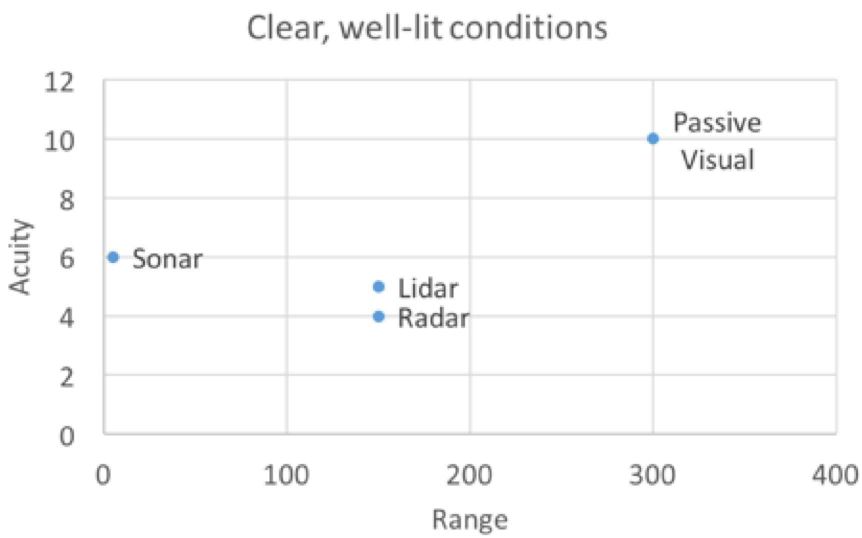
- IMUs are part of an Inertial Navigation System (INS)
  - Use accelerometers and gyroscopes to track position and orientation of an object relative to start position, orientation, and velocity
- There are typically
  - 3 accelerometers measuring linear accelerations
  - 3 orthogonal rate-gyroscopes measuring angular velocities
- Types of IMUs
  - Stable-platform IMUs
    - A platform is used to mount the inertial sensors
    - The platform is isolated from external rotational motion
  - Strapdown IMUs
    - Inertial sensors mounted rigidly

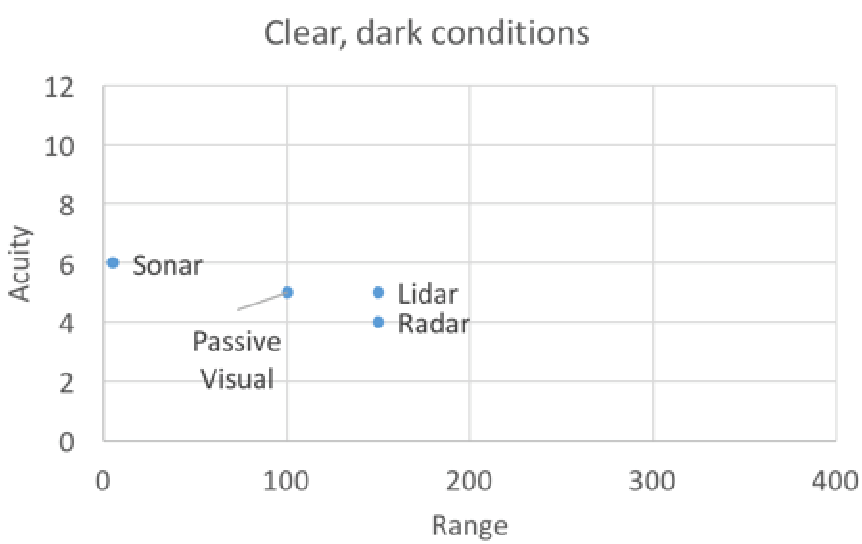
Comparison (1/2)

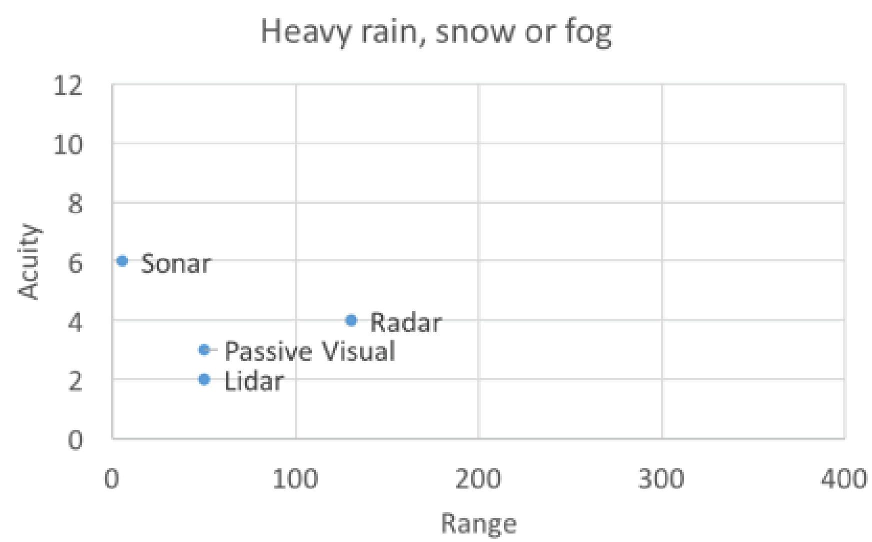


https://cleantechnica.com/2016/07/29/tesla-google-disagree-lidar-right/

Comparison (2/2)



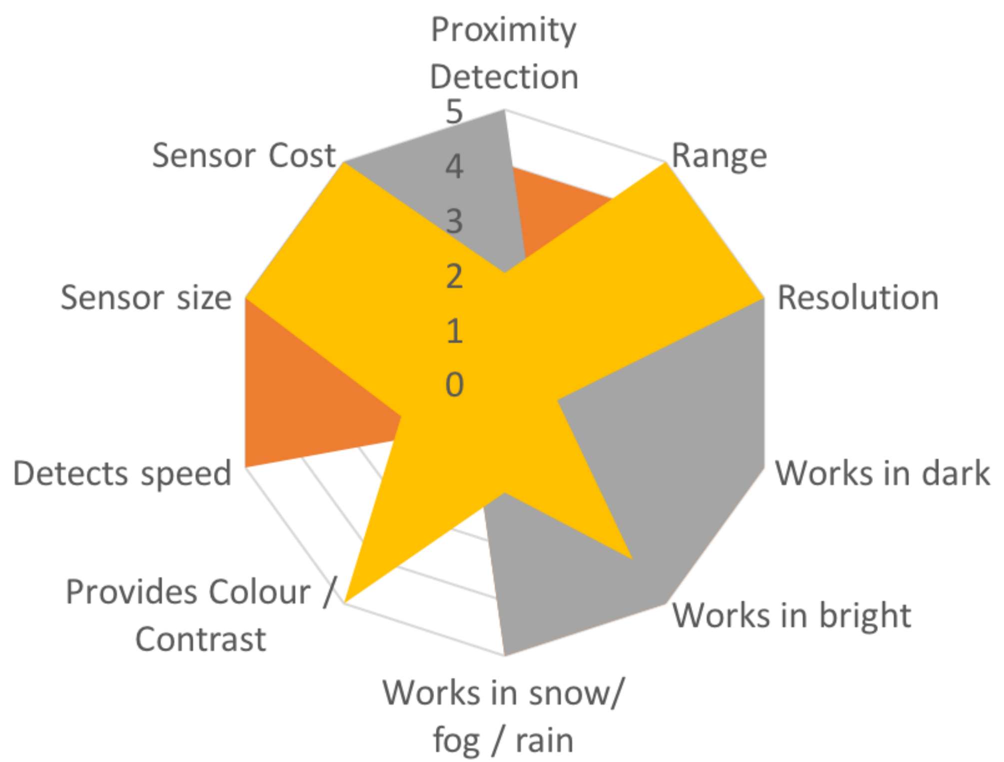




https://cleantechnica.com/2016/07/29/tesla-google-disagree-lidar-right/

Sensor Fusion

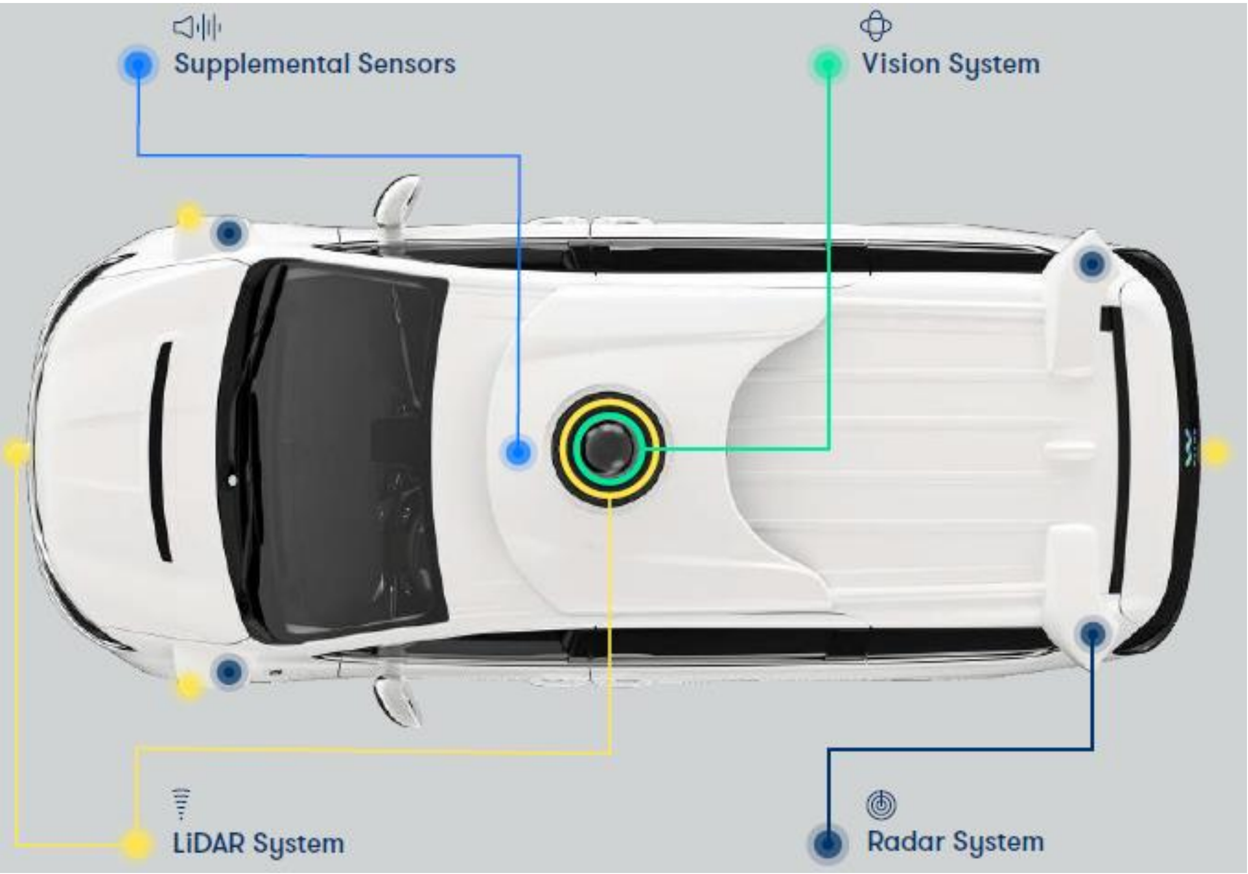
- Combine sensing data from different sources such that the resulting information has less uncertainty [Wikipedia]



https://cleantechnica.com/2016/07/29/tesla-google-disagree-lidar-right/

Waymo's Sensors

- ❑ Lidar, radar, camera (vision), and supplemental sensors
  - ➢ Supplemental sensors: audio sensor and GPS



Waymo Safety Report 2018

Outline

- ❑ Sensing
- ❑ Perception
  - ➢ Perception from Lidar and Camera (Why Lidar and Camera?)
    - **Data Representation**
    - Segmentation Algorithms
    - Object Detection Algorithms
  - ➢ Localization

Data Representation (1/2)

- Following representations for lidar data are popular
  - Point-cloud-based approaches
  - Feature-based approaches
  - Grid-based approaches
- The choice of representation guides the choice of the algorithms for segmentation and detection

- Point-cloud-based approaches
  - Directly use the raw sensor data for further processing
  - Provide a finer representation of the environment, at the expense of increased processing time and reduced memory efficiency
    - A voxel-based filtering mechanism can reduce the number of points

Data Representation (2/2)

Feature-based approaches

- Extract parametric features out of the point cloud and represent the environment using the extracted features
  - The features that are commonly used include lines and surfaces
- Memory-efficient but abstract
  - Its accuracy is subject to the nature of the point cloud, as not all environment features can be approximated well by aforementioned set of feature types

Grid-based approaches

- Discretize the space into small grids, and each of which is filled with information from the point cloud
- Memory-efficient and independent from predefined features
  - However, it is not straightforward to determine the size of the discretization

Outline

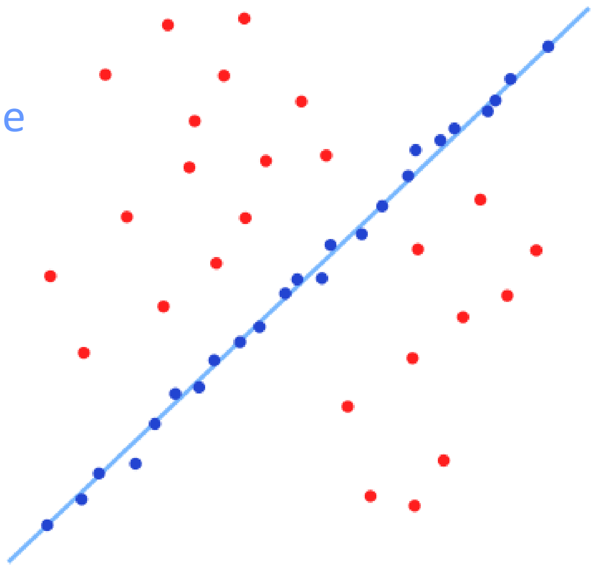
- ❑ Sensing
- ❑ Perception
  - ➢ Perception from Lidar and Camera
    - Data Representation
    - **Segmentation Algorithms**
    - Object Detection Algorithms
  - ➢ Localization

Segmentation Algorithms

- Cluster points into multiple homogenous groups
- Categories
  - Edge-based methods
    - Good when objects have strong artificial edge features (e.g., road curbs)
  - Region-based methods
    - Region-growing: pick seed points and then grow regions based on criteria
  - Model-based methods
    - Fit points into pre-defined categories such as planes, spheres, cones etc.
  - Attribute-based methods
    - First compute attributes for each point and then cluster based on attributes
  - Graph-based methods
    - Cast point cloud into graph-based structures
  - Deep-learning-based methods

Random Sample and Consensus

- Random sample and consensus (RANSAC)
  - Fitting of a model in the presence of outliers
    - Model: line, bounding box, or any parametric shape
  - Algorithm
    - Select n points at random
    - Estimate model values θ by the n points
      - Say the values are θ*, and the resulted shape is S(θ*)
    - Find how many points are within some tolerance of S(θ*)
      - Say this is k
    - If k is large enough, accept the model and exit
    - Repeat several times
  - Parameters matter
    - Pick n based on how many points are required to find a good fit for the shape
    - Pick k based on how many points would lie in the shape



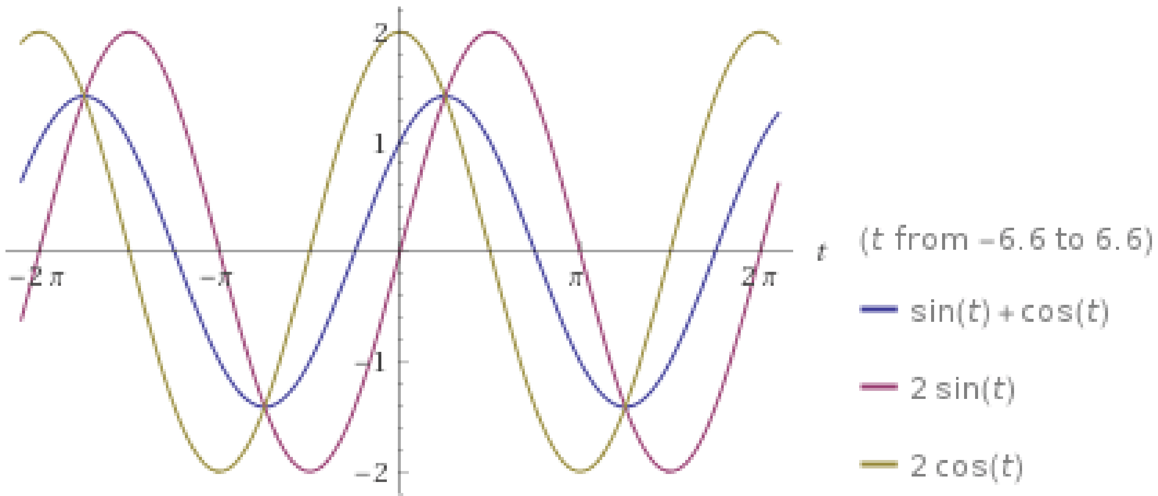
https://en.wikipedia.org/wiki/Random_sample_consensus

Hough Transform (1/3)

- A tool to detect lines, circles, and more general shapes
  - Operates on sets of points and helps obtain a geometric representation of shapes that points may form
- A simple transformation for the concept
  - Map each point in the (x,y) space to a line in the (m,c) space
    - Line in the (m,c) space: $y = mx + c$
  - This allows points to "vote" on which lines best represent them
    - If lines intersect, this represents a collection of points with the slope and intercept defined by the intersection
  - Practice 1: (0,2), (1,1), (2,0)
    - $2 = c$, $1 = m + c$, $0 = 2m + c \rightarrow m = -1$, $c = 2$
  - Practice 2: (0,2), (1,1), (2,0), (0,0)
    - $2 = c$, $1 = m + c$, $0 = 2m + c$, $0 = c \rightarrow m = -1$, $c = 2$

Hough Transform (2/3)

- Map each point in the (x,y) space to a curve in the (r,θ) space
  - Problem with the (m,c) space
    - Vertical line: $m = \infty$
  - Line in the (r,θ) space: $r = x \cos\theta + y \sin\theta$
    - r = length of normal to line
    - θ = angle made by normal with the x-axis
  - Practice 1: (0,2), (1,1), (2,0)
  - Practice 2: (0,2), (1,1), (2,0), (0,0)



Hough Transform (3/3)

- Advantages
  - Conceptually simple and easy to implement
  - Robust to noise
  - Applicable to various shapes beyond lines
  - Handle missing and occluded data gracefully
- Disadvantages
  - Computationally complex if there are many shapes to look for
  - Hard to separate collinear line segments

Outline

- Sensing
- Perception
  - Perception from Lidar and Camera
    - Data Representation
    - Segmentation Algorithms
    - **Object Detection Algorithms**
  - Localization

Object Detection

- Detection from lidar data
  - Segmented clusters are there
  - Traditional machine learning algorithms based on Support Vector Machines (SVMs), Gaussian Mixture Models, etc.
- Detection from camera data
  - Lane-marking detection
  - Drivable path detection
  - On-road object detection
- Image pre-processing before applying detection algorithms
  - Remove obstacles (e.g., other vehicles, raindrops)
  - Weaken shadows
  - Normalize images by controlling camera exposure
  - Limit regions of interest

Lane-Marking Detection

- Common in Advanced Driver-Assist Systems (ADAS)
  - Support Lane Departure Warning System (LDW), Lane Keeping Assistance (LKA), lane change assistance system
- Several decades of work
  - Still not fully solved because of uncertainties in traffic conditions and road-specific issues
    - Shadows, worn-out markings, directional arrows, warning text, pedestrian crossings, etc.
- Common sub-tasks
  - Extract features
  - Fit pixels into various models (lines, parabolas, hyperbolas)
  - Estimate vehicle pose based on the fitted model
  - Use of temporal continuity
  - Transform to world coordinates

Drivable Path Detection

- Detect road boundaries where vehicles can drive freely and legally without collisions
- Deep-learning-based methods seem to be effective
- Other algorithms include exploiting GPS data and OpenStreetMap data

On-Road Object Detection

- Detect other vehicles, pedestrians, bicycles, etc.
- Deep-learning-based methods seem to be effective
- General pipeline for deep learning approaches
  - Generate a set of proposal bounding boxes in the input image
  - Each proposal box is passed through a Convolutional Neural Network (CNN) to obtain a label and fine tune the bounding boxes

- NVIDIA object detection
  - https://www.youtube.com/watch?v=KS_4xjXNTxg

Outline

- Sensing
- Perception
  - Perception from Lidar and Camera
    - Data Representation
    - Segmentation Algorithms
    - Object Detection
  - **Localization**

Localization

- Most common approach is to combine GPS, vehicle odometry, and Kalman filter
  - This becomes unreliable when GPS signal quality is poor
    - Example: urban environments, tunnels, tall buildings, etc.
- Map-aided localization
  - Use local features to achieve precise localization
  - Simultaneous Localization And Mapping (SLAM)

SLAM

- Motion model
  - Capture the motion of a vehicle
  - May be inaccurate because of actuation errors
- Inverse observation model
  - Determine the positions of landmarks (usually not in the map) from observation by mathematical models
  - May be inaccurate because of sensing errors
- Direct observation model
  - Observe previously mapped landmarks and use them to correct the self-localization and the positions of landmarks in the map
- SLAM = three models and an estimator
- MATLAB SLAM
  - https://www.youtube.com/watch?v=XZxpmS0QuHI

Q&A