Computer Vision for Autonomous Systems

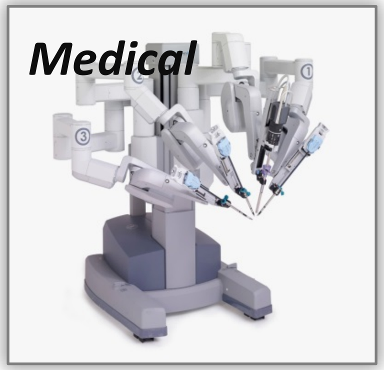
Minh N. Do

ADSC
Illinois at Singapore Pte Ltd

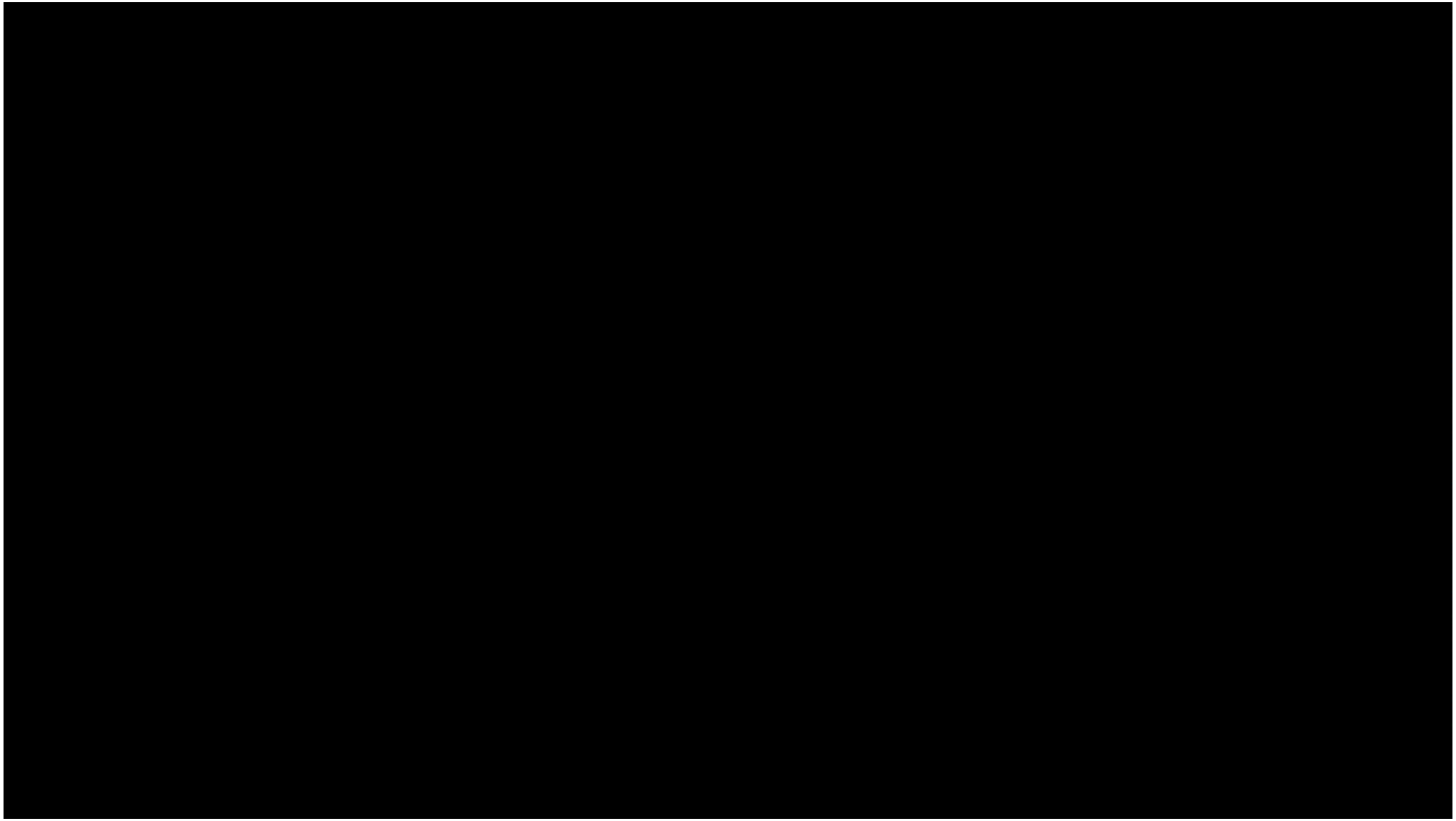
ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Autonomous Systems (AS)

- Automobile
- Service
- Consumer
- Medical
- Entertainment
- Education
- Domestic
- Manufacturing
- Military
- Augmented Reality



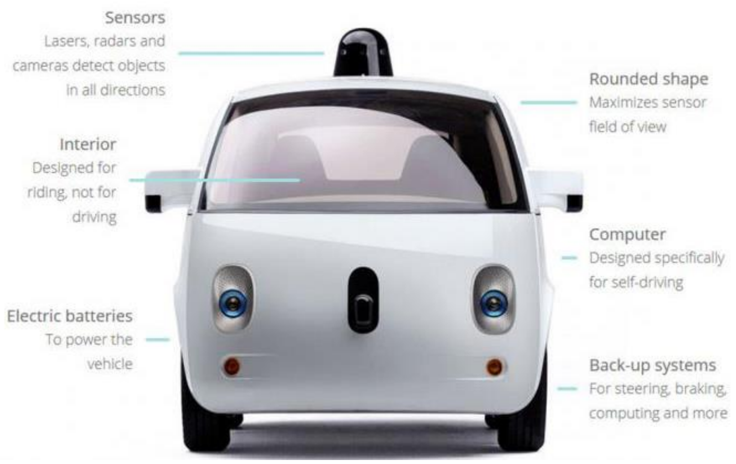
Example: Toy robot

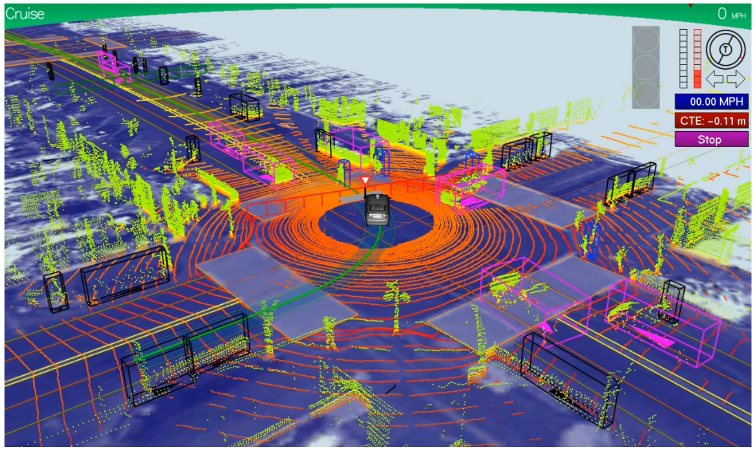


Source: https://petronics.io/

Perceptual capability in dynamic environment

Mapping	Object	Dynamics
• Location • Geometry • Semantics • Updates	• Distance • Dimension • Category • Instance	• Motion • Behavior • Interaction



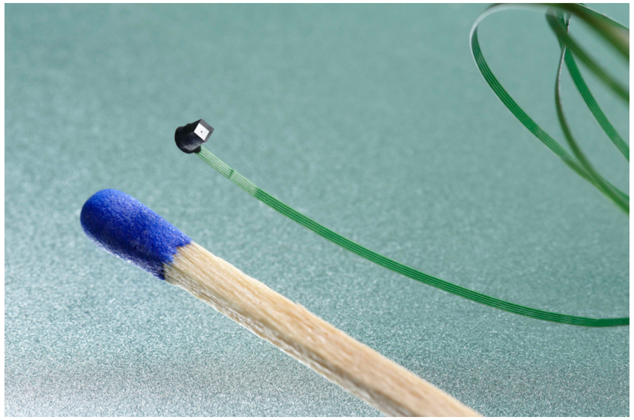


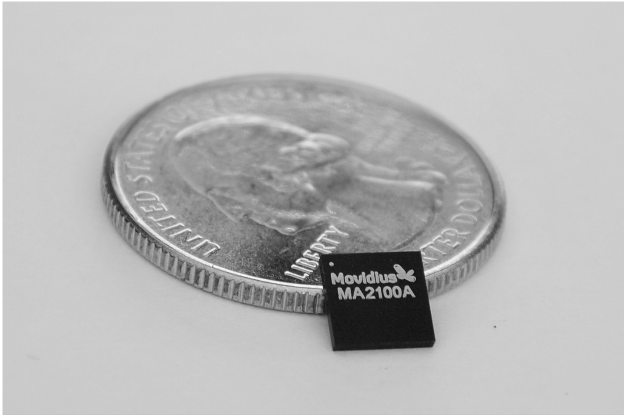
Vision as sensing input

- High resolution provides details about complex scenes
  - State of the art camera has 1.3-1.7 MP, running at 36FPS (47-61M per second)
  - Lidar technology ~60-300k per second

- Shape vs. Appearance
  - Most complex situations are defined by appearance (texture) more than shape:
    - e.g. road markings, traffic signs, person identity, object instance, etc.

- Cheap and versatile in size and configuration





Computer vision

The goal of computer vision is to make computers efficiently perceive, process, and understand visual data such as images and videos. The ultimate goal is for computers to emulate the striking perceptual capability of human eyes and brains-or even to surpass and assist the human in certain ways. – [Microsoft Research]

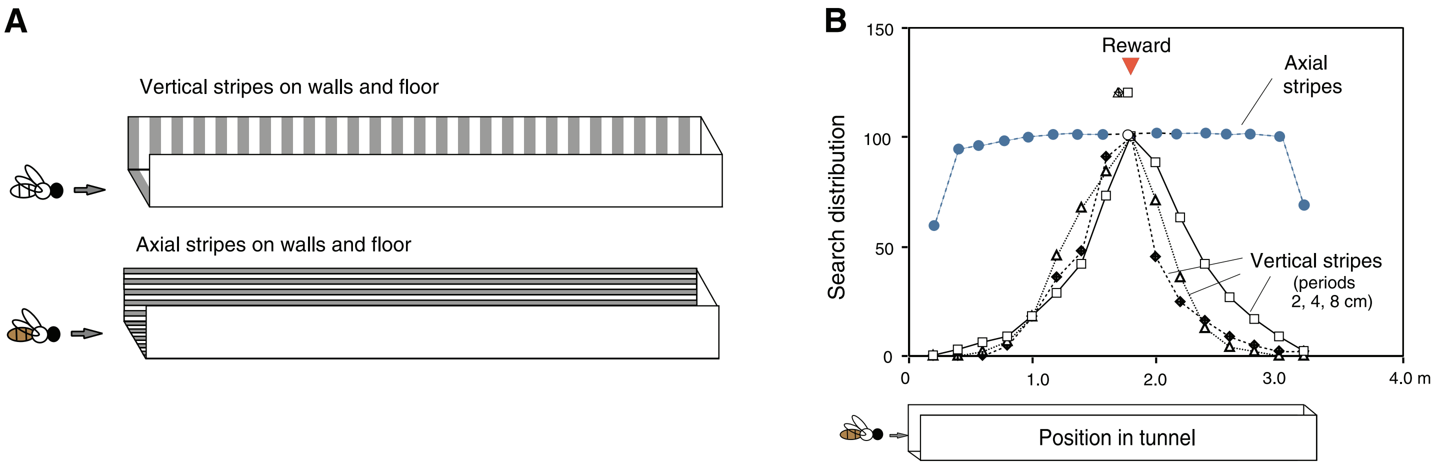
past

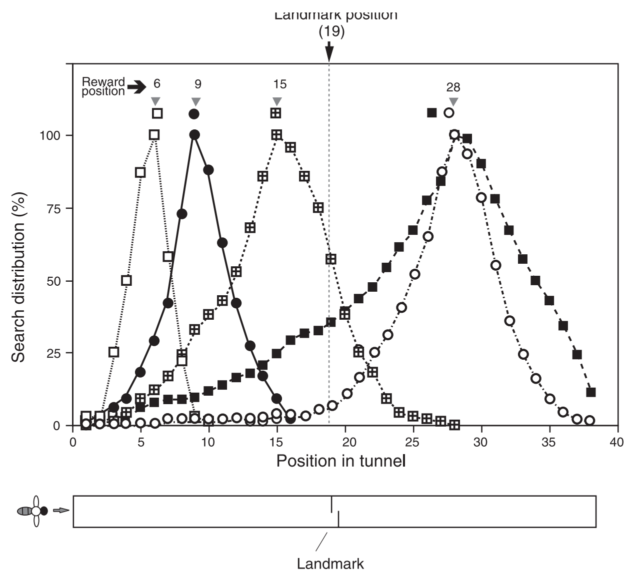
- Single image
- Static scene
- RGB only
- Limited data
- Limited computation power
- Slow algorithms

present

- Video
- Dynamic scene
- Depth, IMU
- Large amount of data
- Visual computing chips
- Real-time algorithms

Lessons from bees

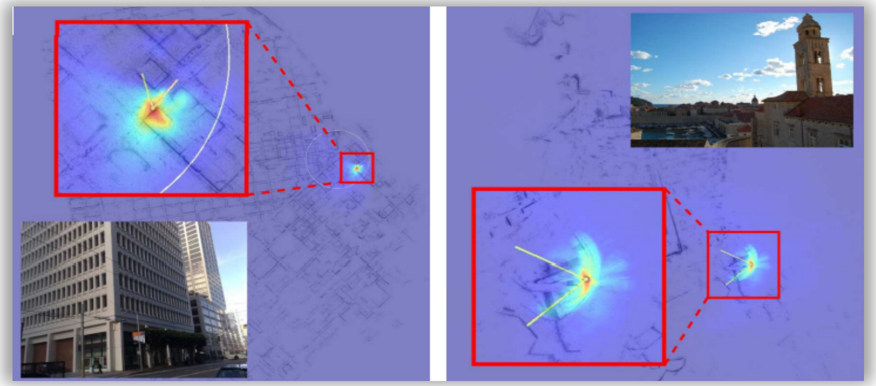




Srinivasan (1997, 2011)

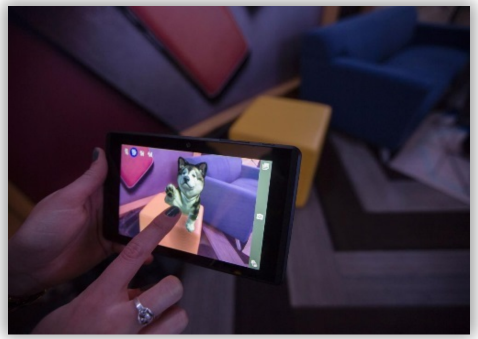
Localization

- Place recognition and localization
- Loop closure detection for SLAM
- Visual SLAM for mobile autonomous system



large scale image-based localization

Map based visual self-localization

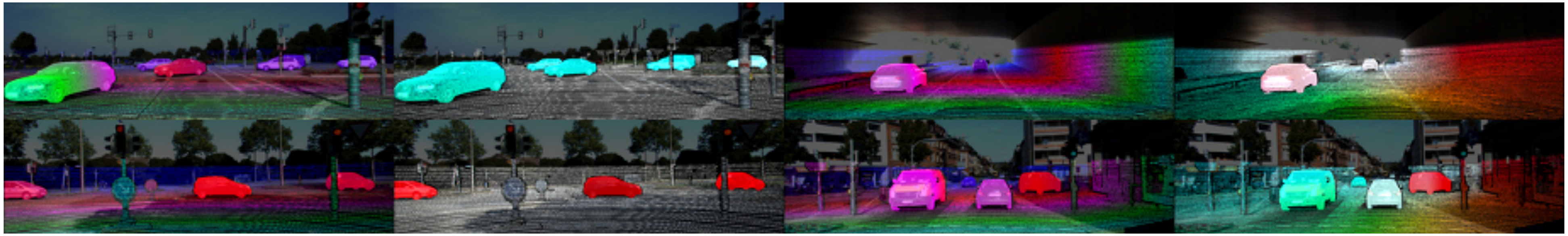


Google Tango

HoloLens

Magic Leap

Depth and motion



3D scene flow

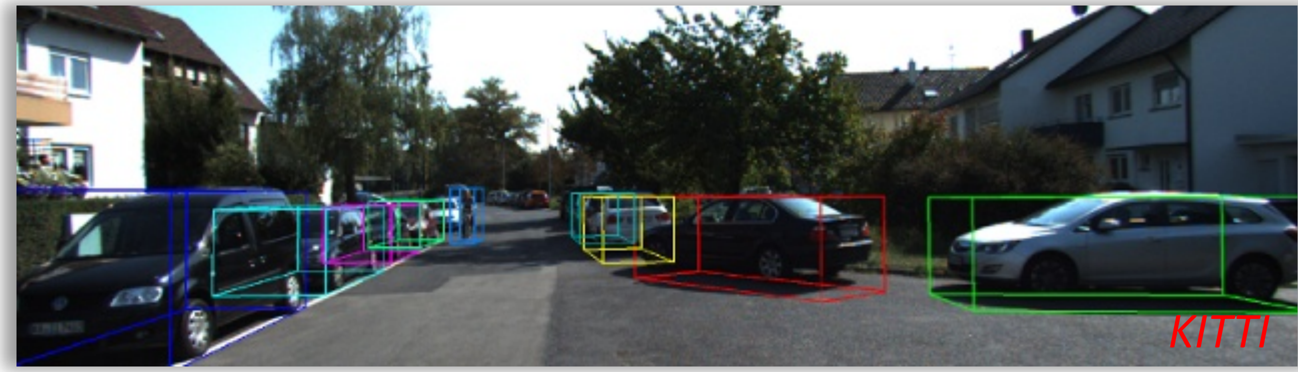
- Per-pixel dense depth and optical flow
- Algorithm complexity and efficiency
- Temporal consistency
- Semantic awareness

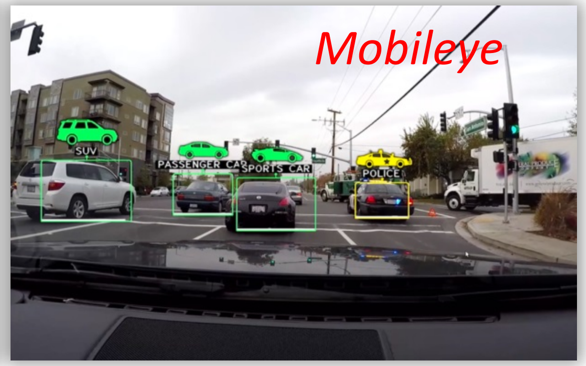
[Bai et al. 2016]

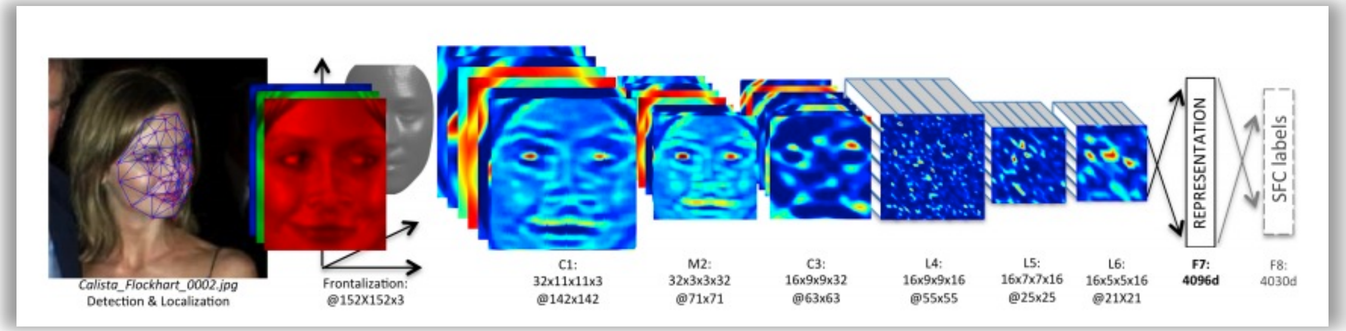
[Kroeger et al. 2016]

Recognizing people, landmarks, and objects

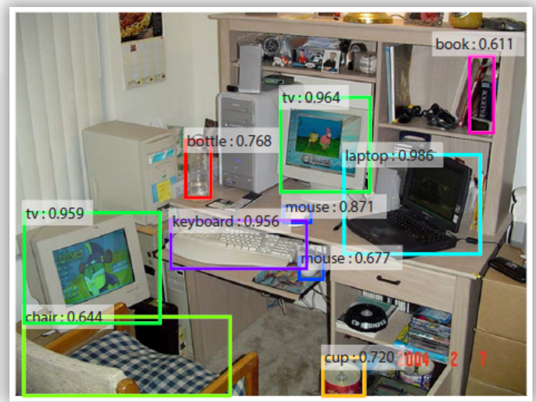
- Detect pedestrians, cars, motorcycles, traffic lights, etc.
- Recognize people and objects







DeepFace 97.25% accuracy vs. human 97.53% accuracy

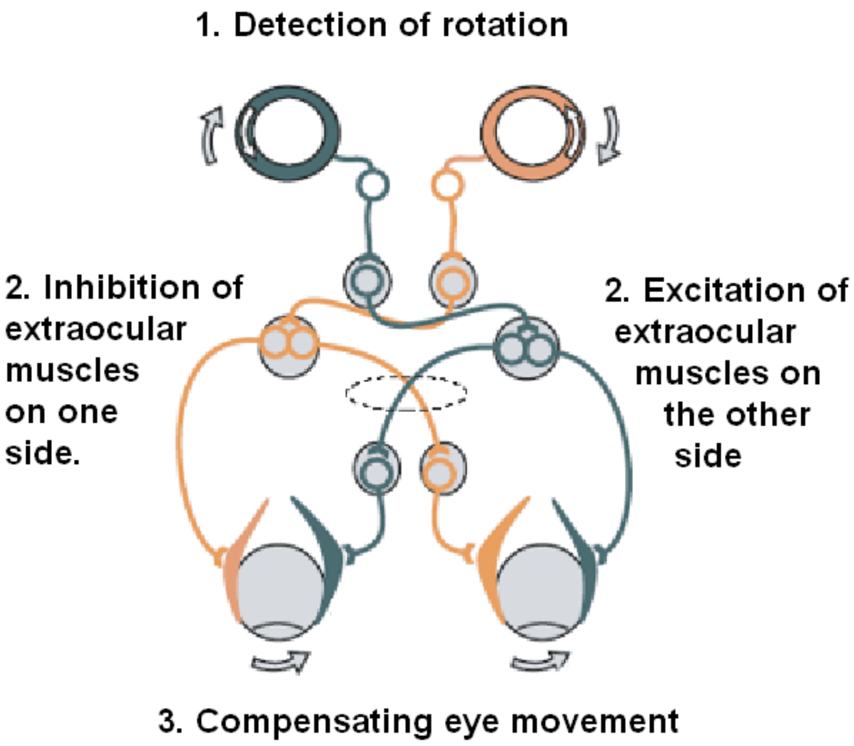


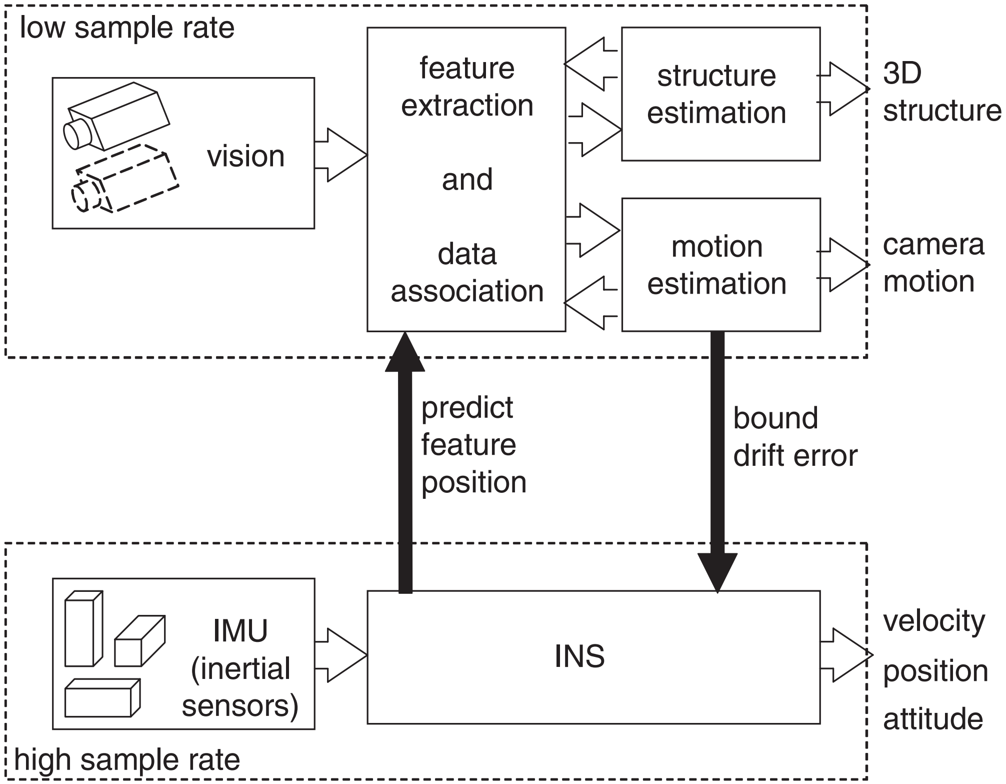
Fast RCNN, 17fps

Perception from a moving platform

Source: Seattle Police Department

Vision + other sensing modality





Summary

- Visual perception is crucial for autonomous systems
  - Small
  - Cheap
  - Fast

- Key problems:
  - Localization and mapping
  - Object and place recognition
  - Motion and dynamics

- Adding other sensing modalities (depth, IMU) significantly helps vision