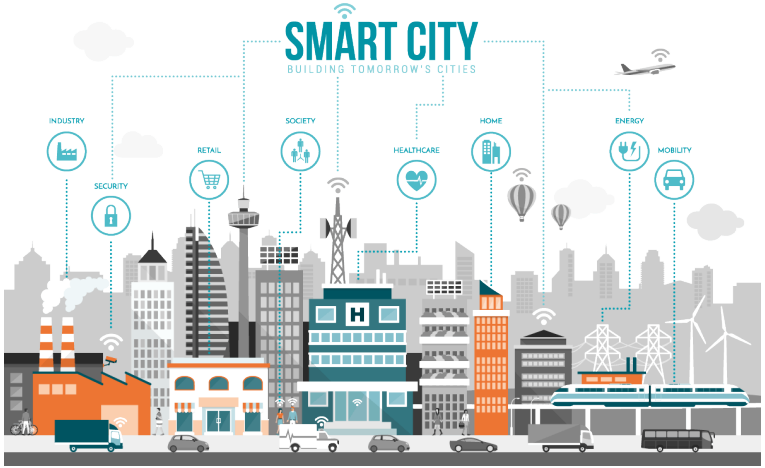
UCLA Samueli School of Engineering

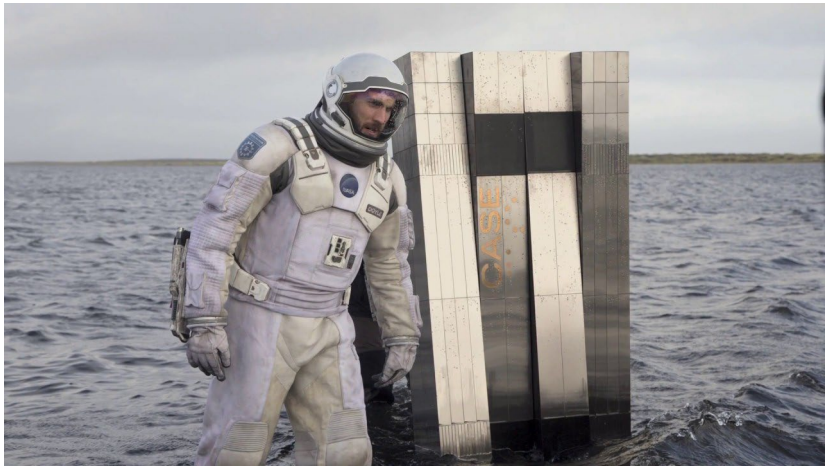
Capricorn: Towards Real-time Rich Scene Analysis Using RF-Vision Sensor Fusion

Ziqi Wang, Ankur Sarker, Jason Wu, Derek Hua, Gaofeng Dong, Akash Deep Singh, and Mani B. Srivastava
University of California, Los Angeles, Electrical and Computer Engineering Department

Real-time Rich Scene Analysis

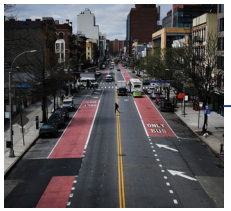
- Situation awareness (SA) is crucial to smart systems
- SA requires the collection and analysis of sensory data to estimate the states of an environment



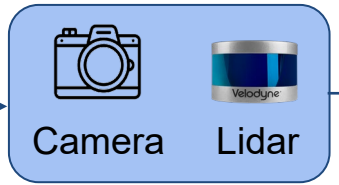


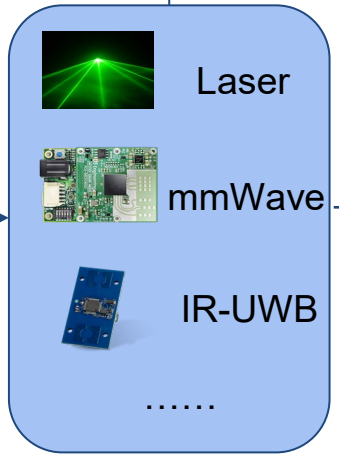
https://www.arcweb.com/industries/smart-cities
Interstellar: https://en.wikipedia.org/wiki/Interstellar_(film)

Real-time Rich Scene Analysis



Scene





Extrinsic Properties

Object Location and Type

Object Activity Status

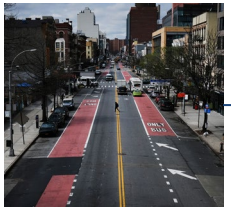
Intrinsic Properties

Examples
Object Type
Shape
Color
Location …

Semantic Labelling

Examples
Human:
Health/ Emotions
Washing Machine:
Wash/ Dry/ Idle
Fan:
Speed Settings
...

Real-time Rich Scene Analysis



Scene

Camera

Lidar

Laser

mmWave

IR-UWB

……

Extrinsic Properties

Object Location and Type

Examples
Object Type
Shape
Color
Location …

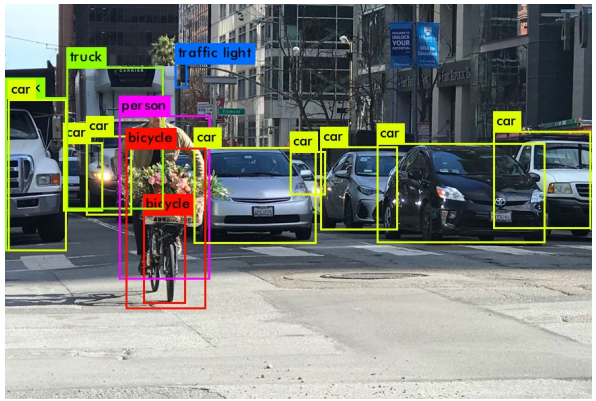
Semantic Labelling

Object Activity Status

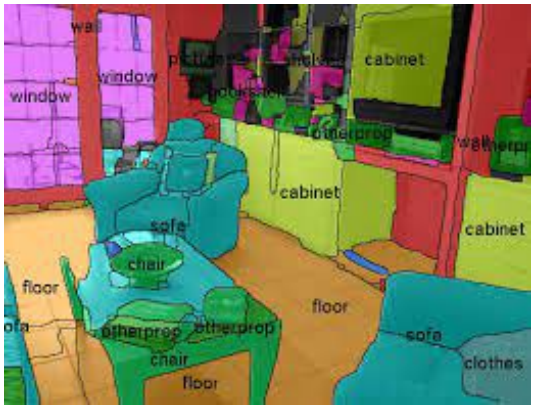
Intrinsic Properties

Examples
Human:
Health/ Emotions
Washing Machine:
Wash/ Dry/ Idle
Fan:
Speed Settings
...

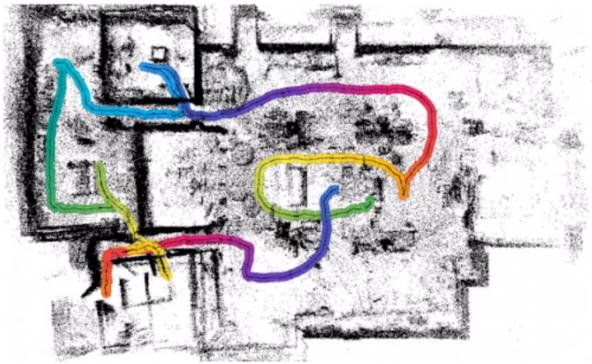
Computer Vision for Extrinsic States



Object Detection and Recognition

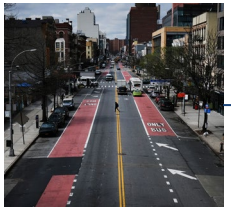


Semantic Segmentation



Localization and Mapping

Real-time Rich Scene Analysis



Scene

Camera

Lidar

Laser

mmWave

IR-UWB

……

Extrinsic Properties

Object Location and Type

Intrinsic Properties

Object Activity Status

Examples
Object Type
Shape
Color
Location …

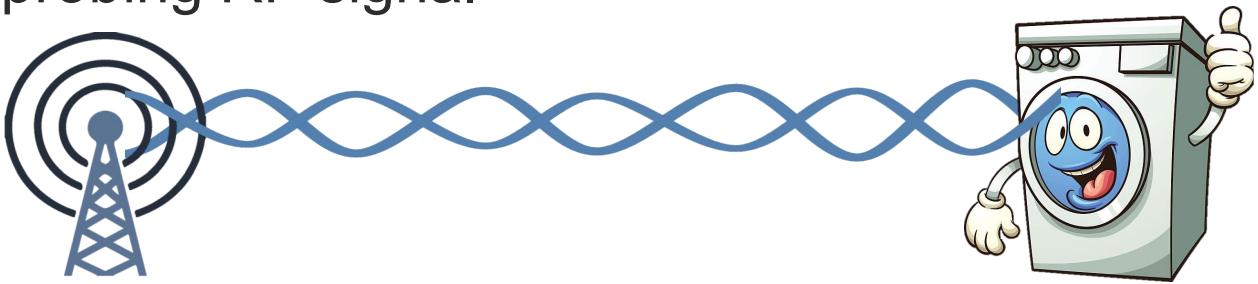
Semantic Labelling

Examples
Human:
Health/ Emotions
Washing Machine:
Wash/ Dry/ Idle
Fan:
Speed Settings
…

Wireless Vibrometry for Intrinsic States

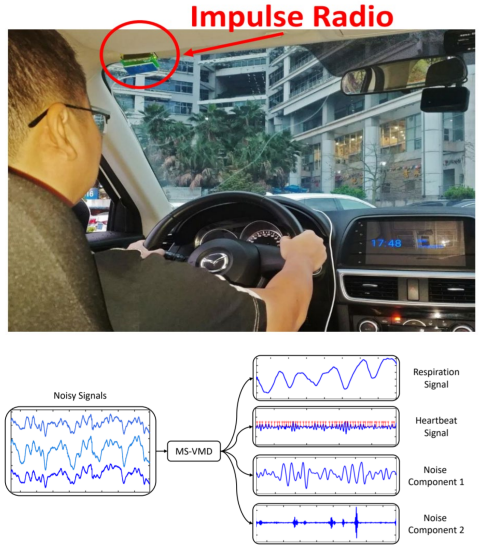
Signal reflect on the surface
Modulated by vibration

Emit a probing RF signal

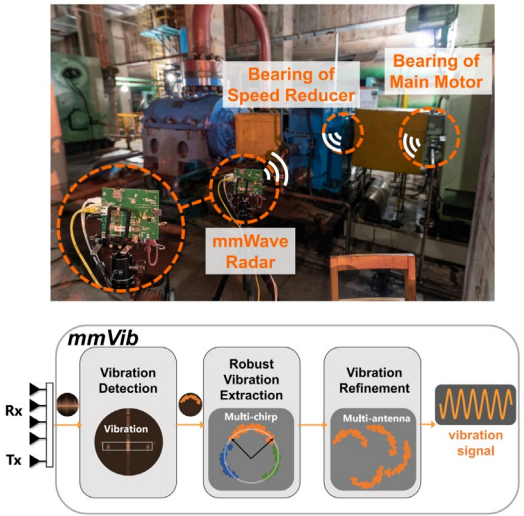


Washing machine
State: washing

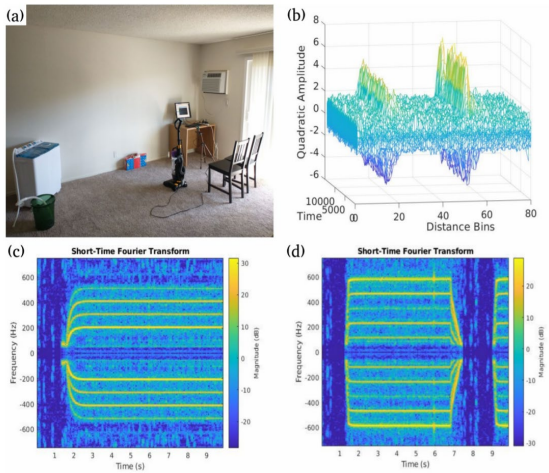
Wireless Vibrometry for Intrinsic States



Vital Signals / Activities from Human
(Vi2Fi, IMWUT'20)



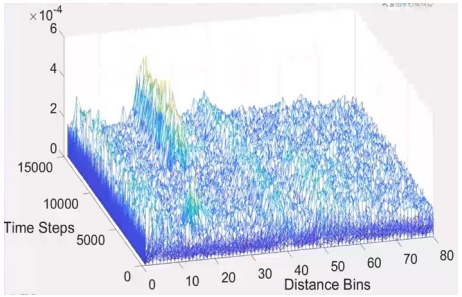
Operating States of Industrial Machinery
(mmVib, MobiCom'20)

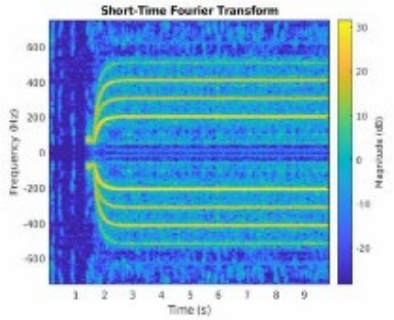


Usage of Home Appliances
(UWHear, SenSys'20)

RF Signals Require Context to Make Sense

- Presumed object type: existing wireless vibrometry sensing systems presuppose the existence of a particular type of object (e.g., a person) in the scene to process the signal accordingly.

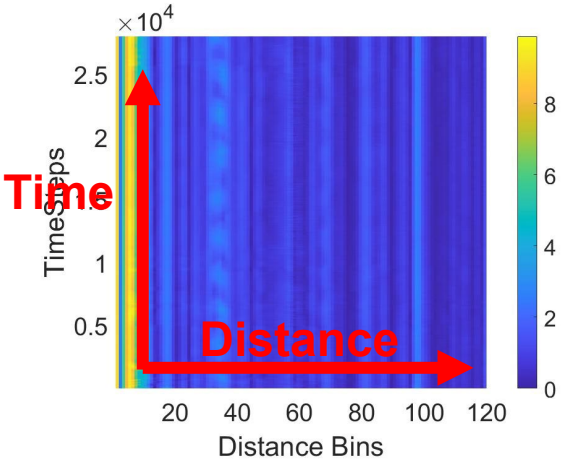




RF Signals Require Context to Make Sense

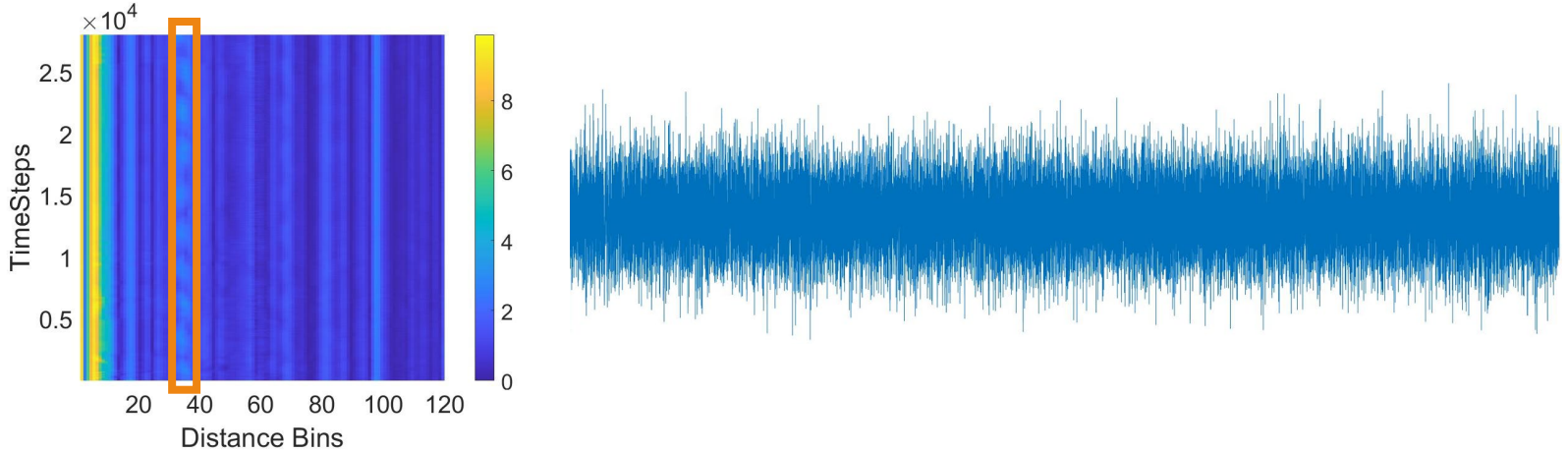
- Presumed object type: existing wireless vibrometry sensing systems presuppose the existence of a particular type of object (e.g., a person) in the scene to process the signal accordingly.

- Presumed number of objects: existing systems either target at a fixed number of subjects or rely on threshold-based search algorithms

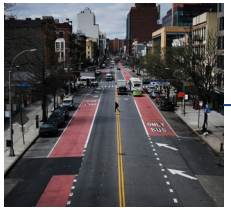


RF Signals Require Context to Make Sense

- Presumed object type: existing wireless vibrometry sensing systems presuppose the existence of a particular type of object (e.g., a person) in the scene to process the signal accordingly.
- Presumed number of objects: existing systems either target at a fixed number of subjects or rely on threshold-based search algorithms



Our Vision: Multimodal Sensor Fusion



Scene

Camera

Lidar

Laser

mmWave

IR-UWB

……

Extrinsic Properties

Object Location and Type

Object Activity Status

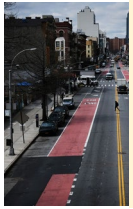
Intrinsic Properties

Examples
Object Type
Shape
Color
Location …

Semantic Labelling

Examples
Human:
Health/ Emotions
Washing Machine:
Wash/ Dry/ Idle
Fan:
Speed Settings
...

Our Vision: Multimodal Sensor Fusion



Multiple sensing modalities can offer complimentary capabilities

Scene

Camera

Lidar

mmWave

IR-UWB

……

Extrinsic Properties

Object Location and Type

Object Activity Status

Intrinsic Properties

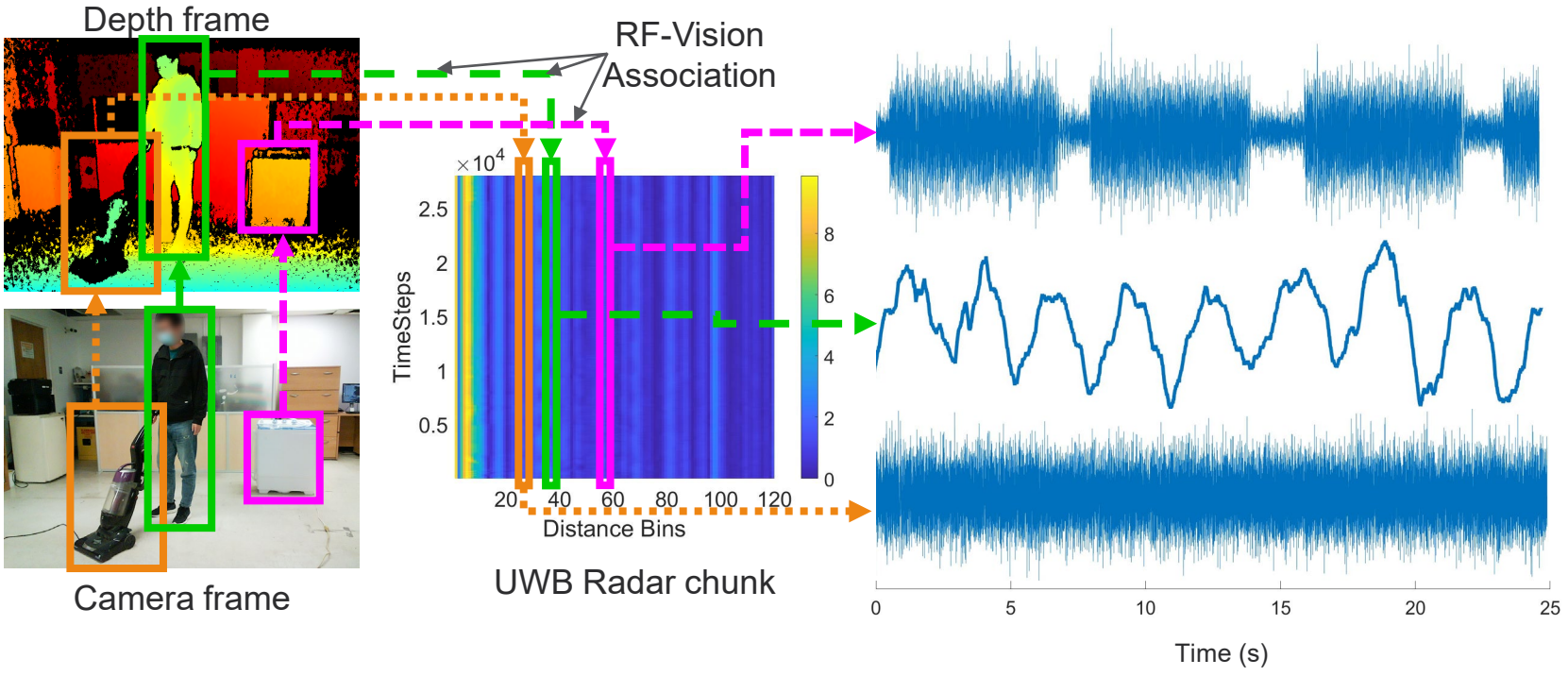
Examples
Object Type
Shape
Color

Human:
Health/ Emotions
Washing Machine:
Wash/ Dry/ Idle
Fan:
Speed Settings
...

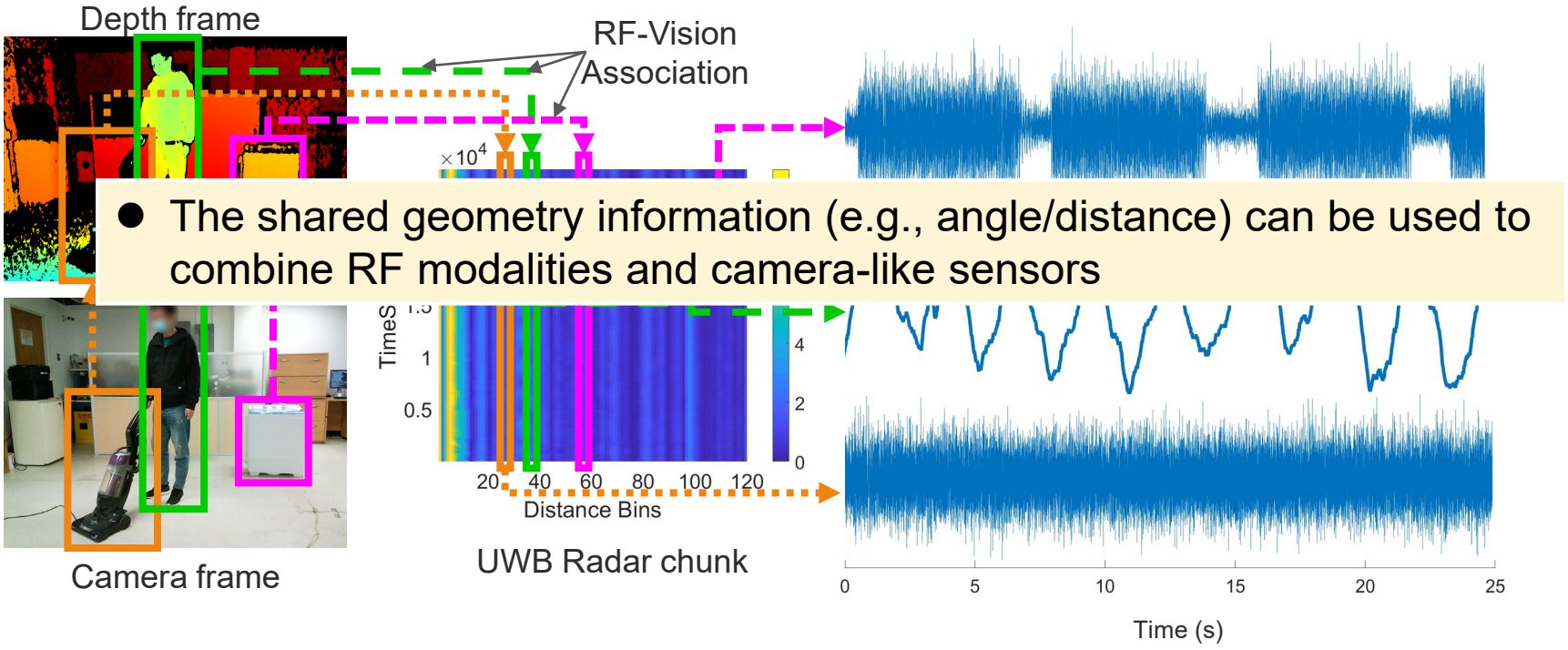
Our Insights: Multimodal Sensor Fusion

- Presumed object type: existing wireless vibrometry sensing systems presuppose the existence of a particular type of object (e.g., a person) in the scene to process the signal accordingly.
- Presumed number of objects: existing systems either target at a fixed number of subjects or rely on threshold-based search algorithms
- Multiple sensing modalities can offer complimentary capabilities
- The shared geometry information (e.g., angle/distance) can be used to combine RF modalities and camera-like sensors
- One sensor's inference can become the prior information for another and expedite the signal processing algorithms

Capricorn Design

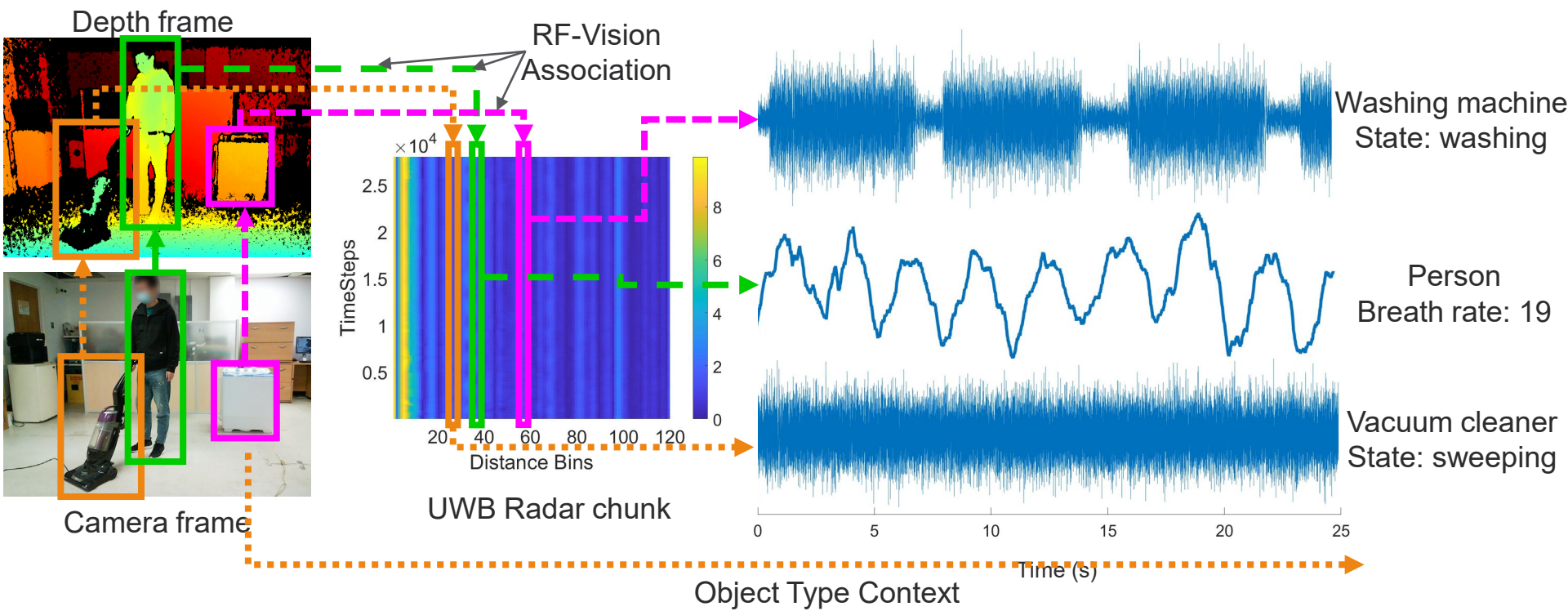


Capricorn Design

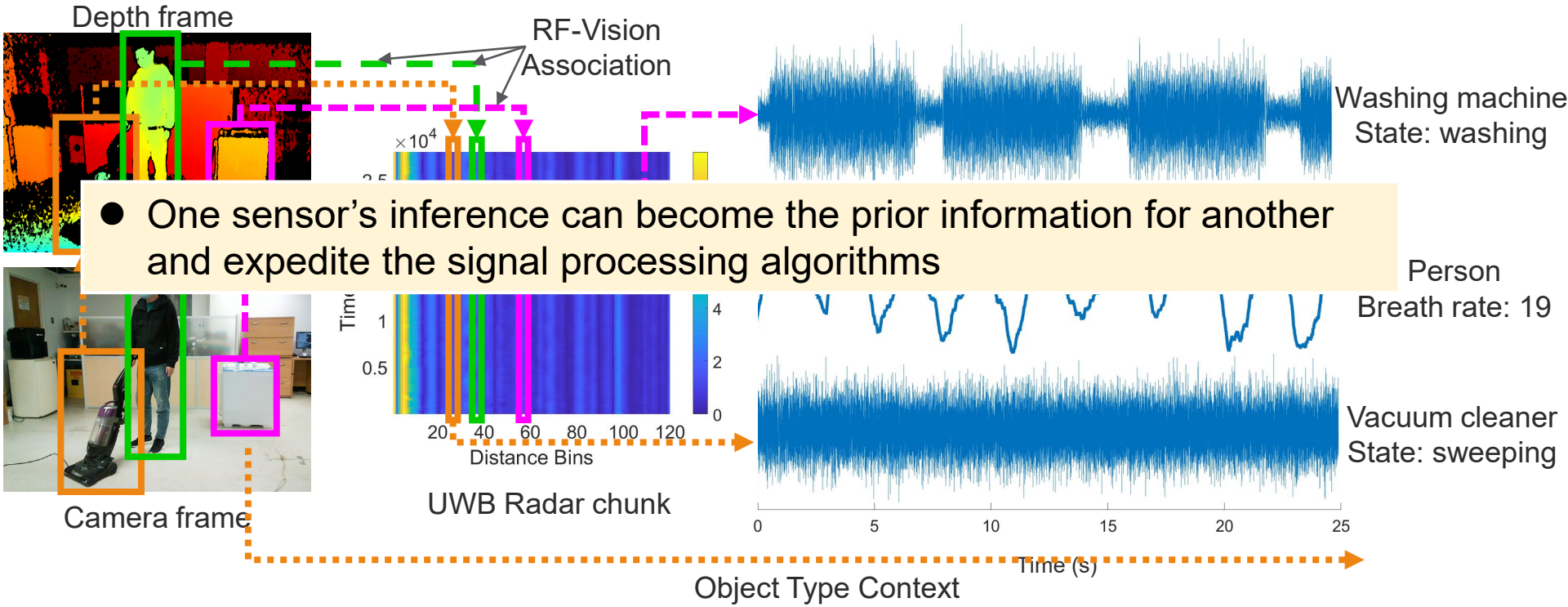


- The shared geometry information (e.g., angle/distance) can be used to combine RF modalities and camera-like sensors

Capricorn Design

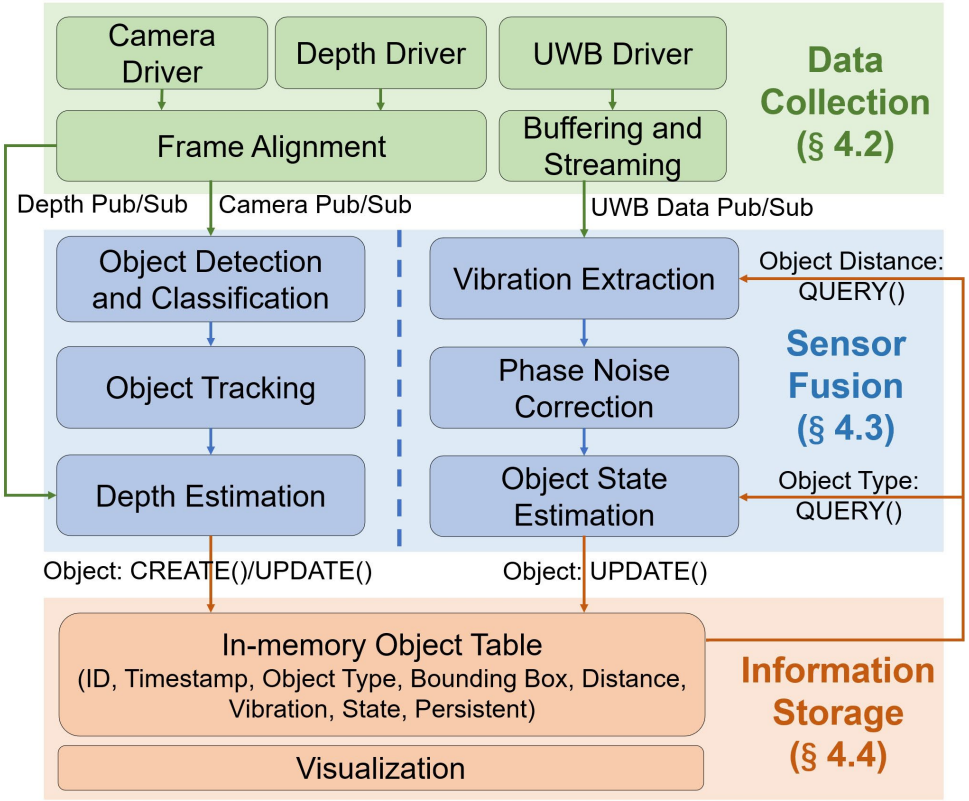


Capricorn Design

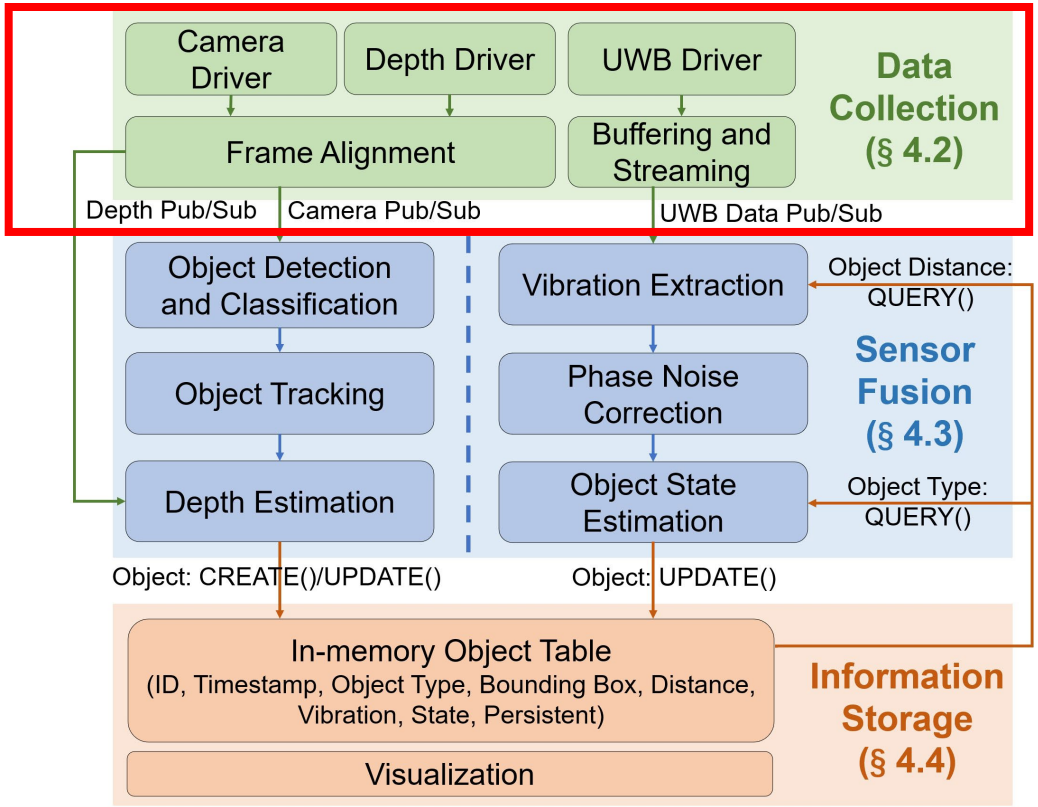


- One sensor's inference can become the prior information for another and expedite the signal processing algorithms

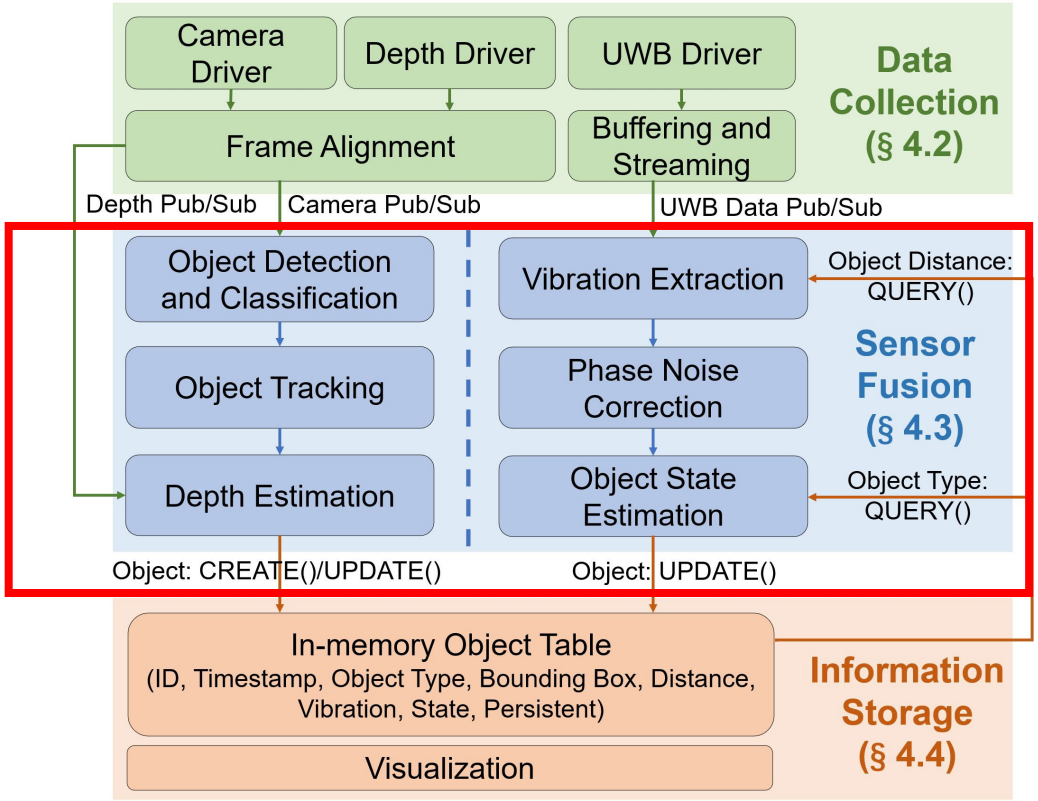
Capricorn Architecture



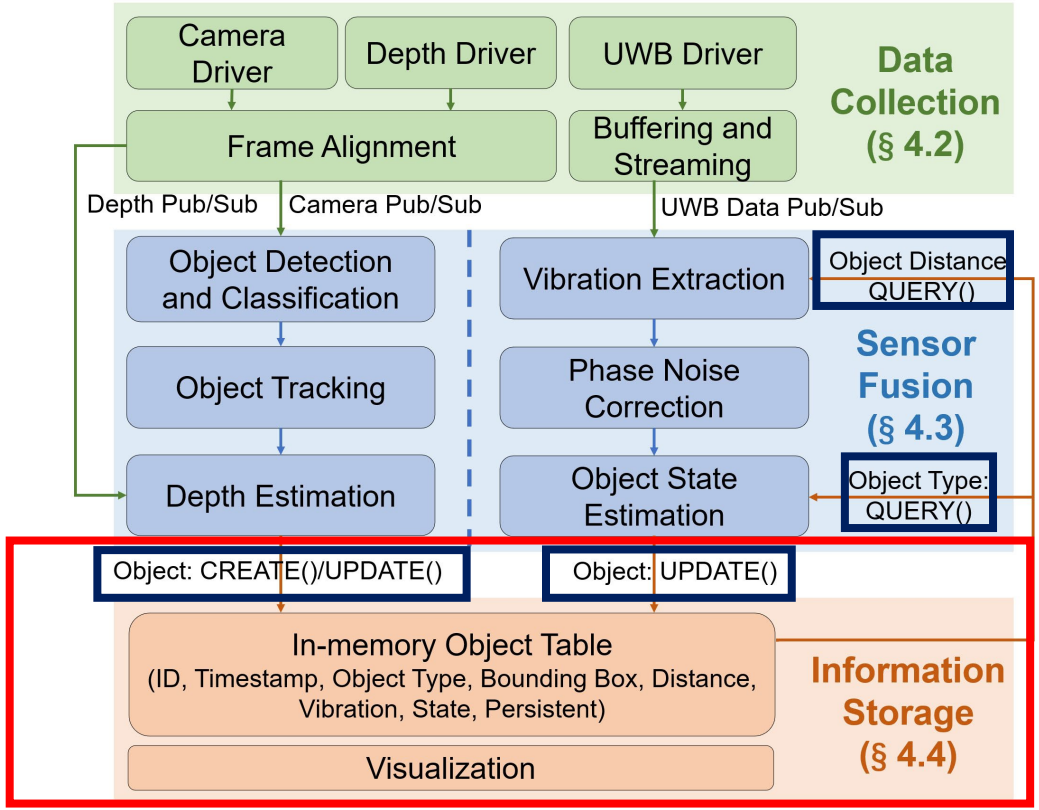
Capricorn Architecture



Capricorn Architecture

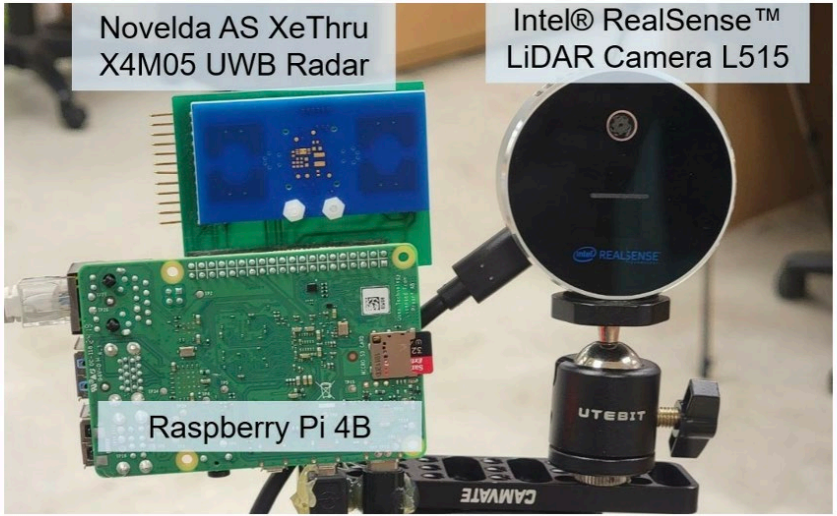


Capricorn Architecture



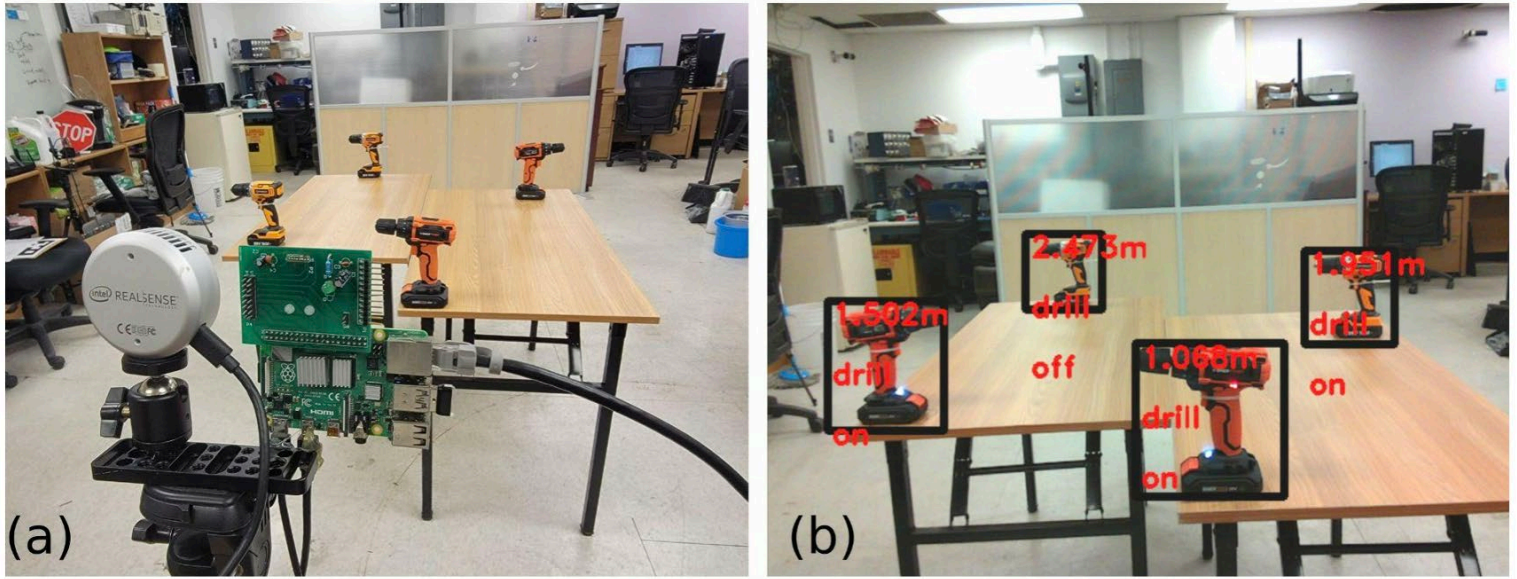
Implementations

- UWB Radar: Xethru X4M05
- LiDAR-Camera: Intel RealSense L515
- UWB Radar Host: Raspberry Pi 4B
  - Cortex-A72 Processor
  - 8 GB RAM
- Main Computation: Intel NUC
  - Intel i7-6770HQ CPU
  - 16 GB RAM
  - No GPU or any hardware accelerator



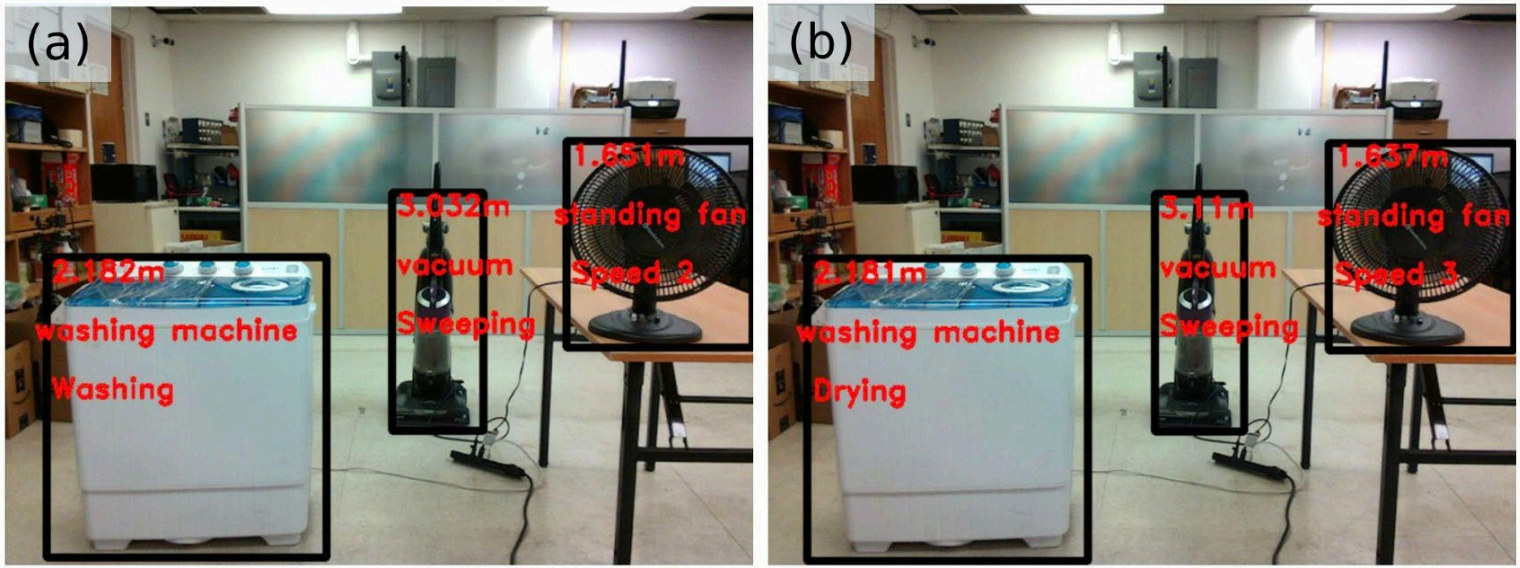
Hardware platform

Sample Applications: It's all real-time!

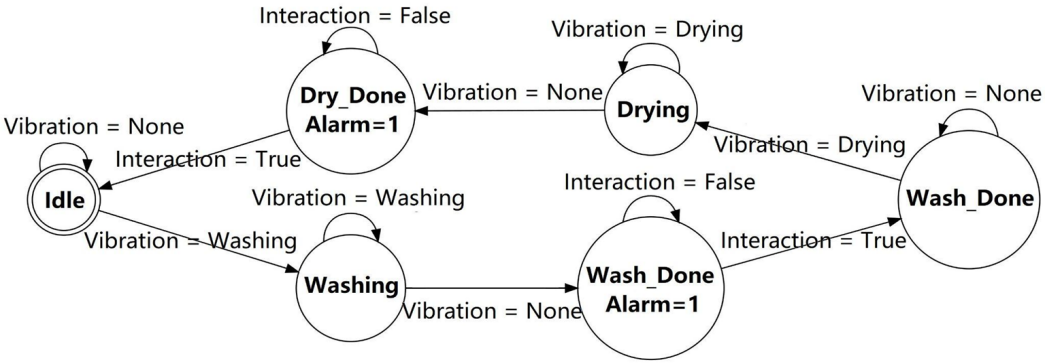


Workshop scene: drill state detection

Sample Applications: It's all real-time!

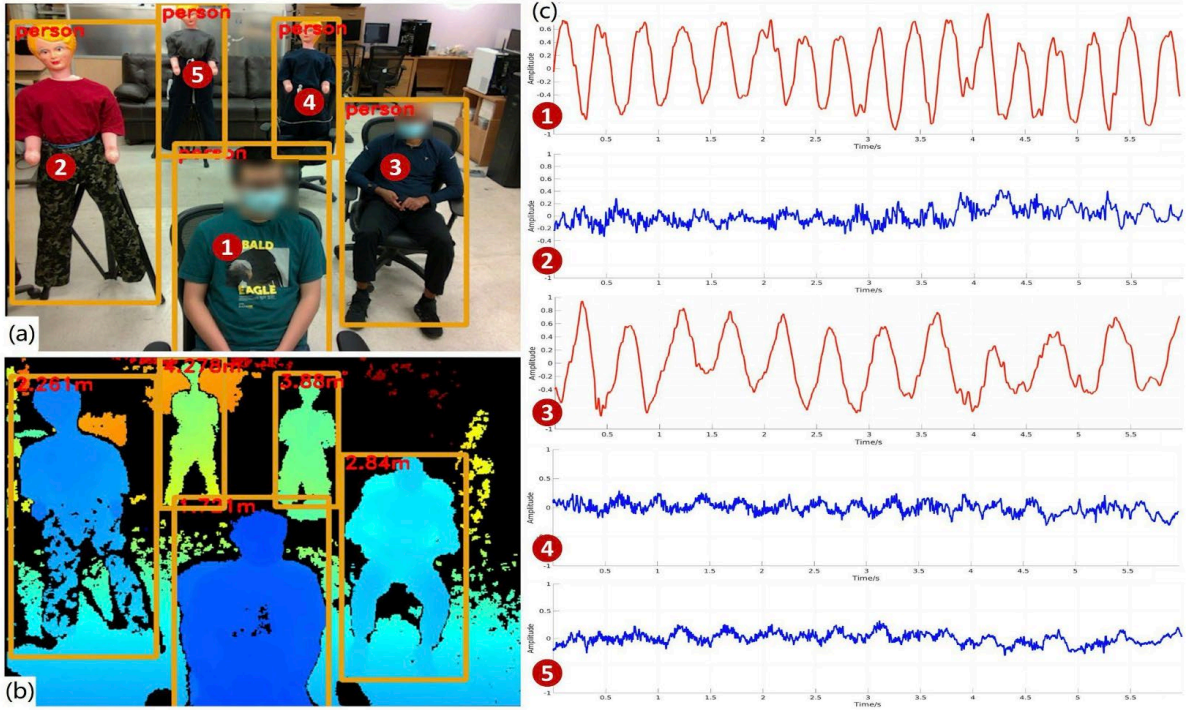


SmartHome scene: appliance usage tracking



Provide a **richer** set of **atomic events** for complex event detection

Sample Applications: It's all real-time!



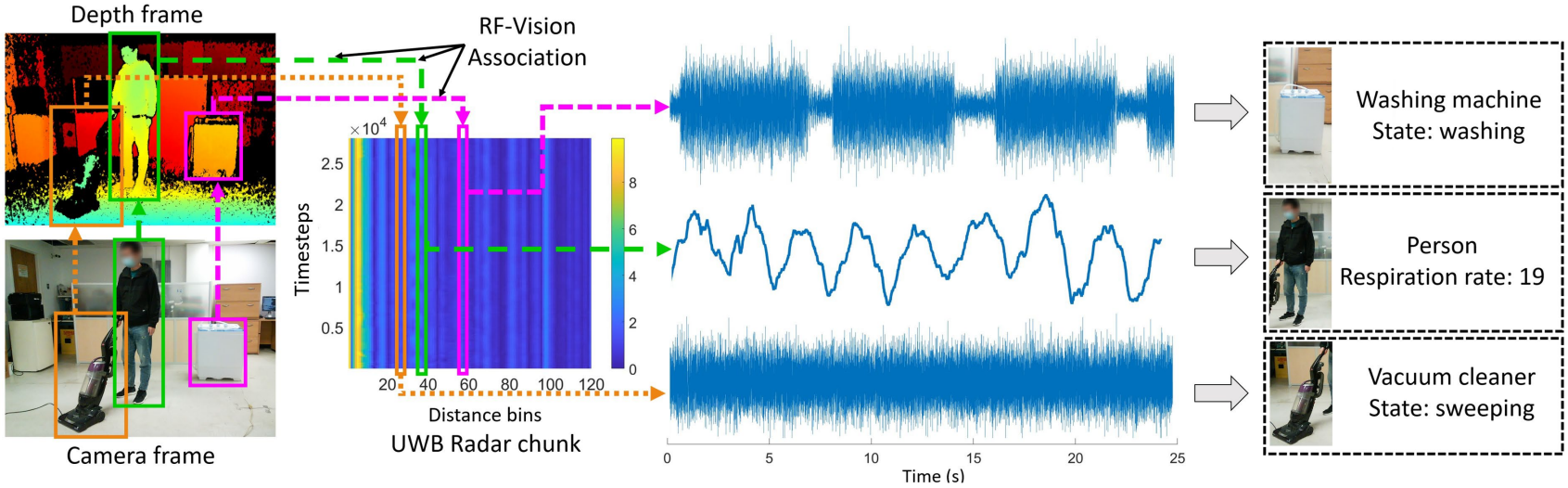
Multi-person respiration rate estimation

Limitations and Future Work

- **Integrate More Sensing Modalities:**
  - Currently we integrate a LiDAR camera and a UWB Radar
  - More sensing modalities will bring in richer information about the scene
  - e.g., Thermal camera, mmWave Radar, ……

- **Enable Mobility:**
  - Current system is mounted statically on a tripod
  - Mobility will give us more perspectives and a better scene understanding
  - SLAM on robots will enable the system to explore a totally new environment

- **Applications:**
  - Integrate sensor fusion information into AR/VR platform
  - Create an enhanced “digital twin” between the physical and virtual world
  - Neural symbolic-based complex event detection



Thank you!

UCLA NESL Networked & Embedded Systems Laboratory

Code

Demo

ARL IoBT REIGN

NIH mDOT

NSF

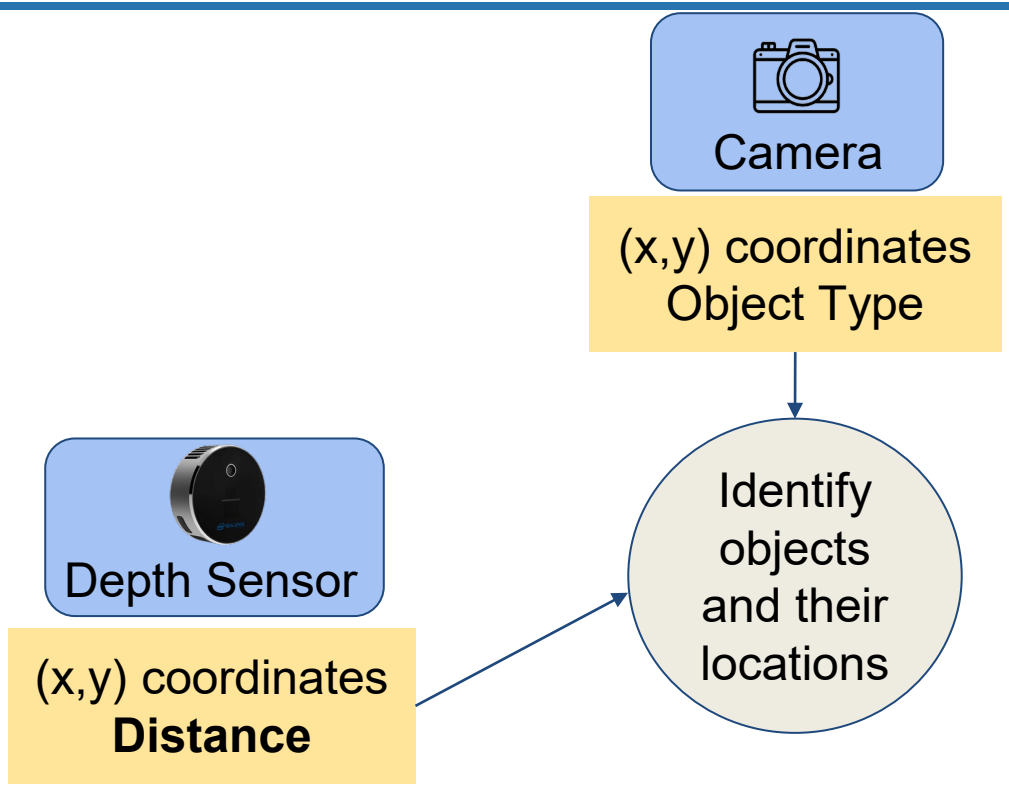
SRC JUMP

CONIX

UCLA Samueli School of Engineering

Backup Slides

Our Vision: Multimodal Sensor Fusion



Our Vision: Multimodal Sensor Fusion

Camera

(x,y) coordinates
Object Type

Depth Sensor

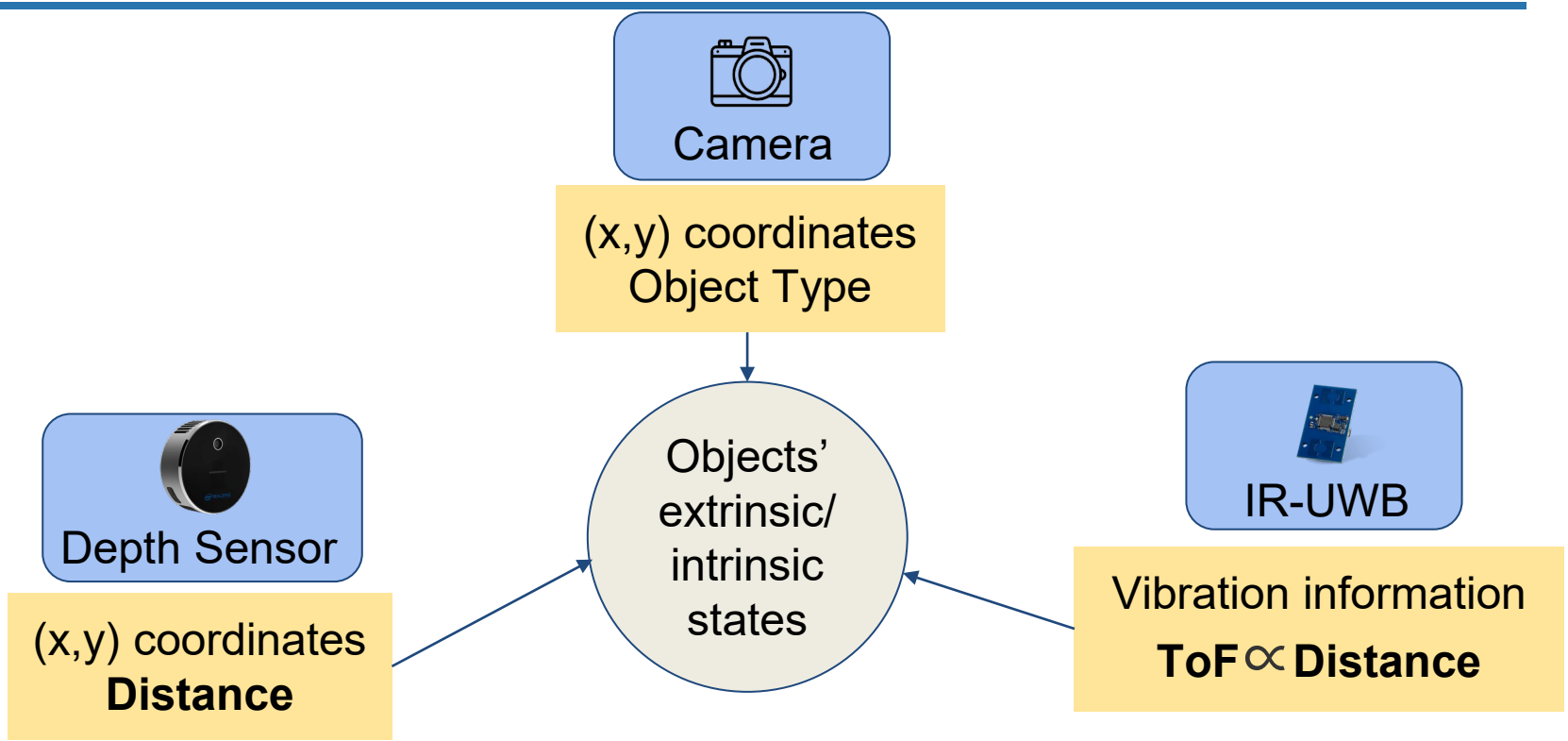
(x,y) coordinates
Distance

RF-Vision Association

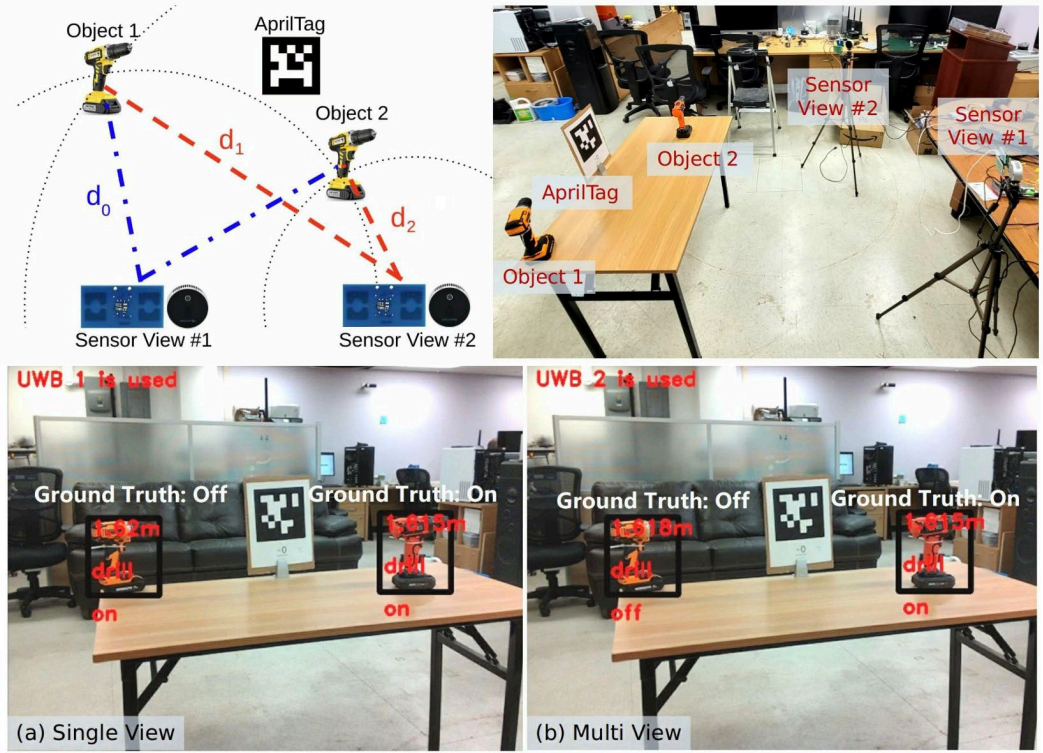
IR-UWB

Vibration information
ToF ∝ Distance

Our Vision: Multimodal Sensor Fusion



Multi-view Capricorn



Multi-view version to distinguish objects at the same distance

Back Up Slides: Latency

Capricorn Component	Mean(ms)	Std(ms)
Camera/Depth Pub-Sub Delay	1.08	0.13
YOLOv5	38.35	5.25
YOLOv5 (GPU)	6.28	1.29
Whole Extrinsic Sensing Pipeline	42.81	6.3
UWB Chunk Pub-Sub Delay	171.61	21.87

Table 2: Latency analysis of Capricorn in the appliance usage classification scene.