Computer Vision in Robotics

CS 6384 Computer Vision
Professor Yu Xiang
The University of Texas at Dallas

What is a Robot?

- A robot is a machine capable of carrying out a complex series of actions automatically (Wikipedia)

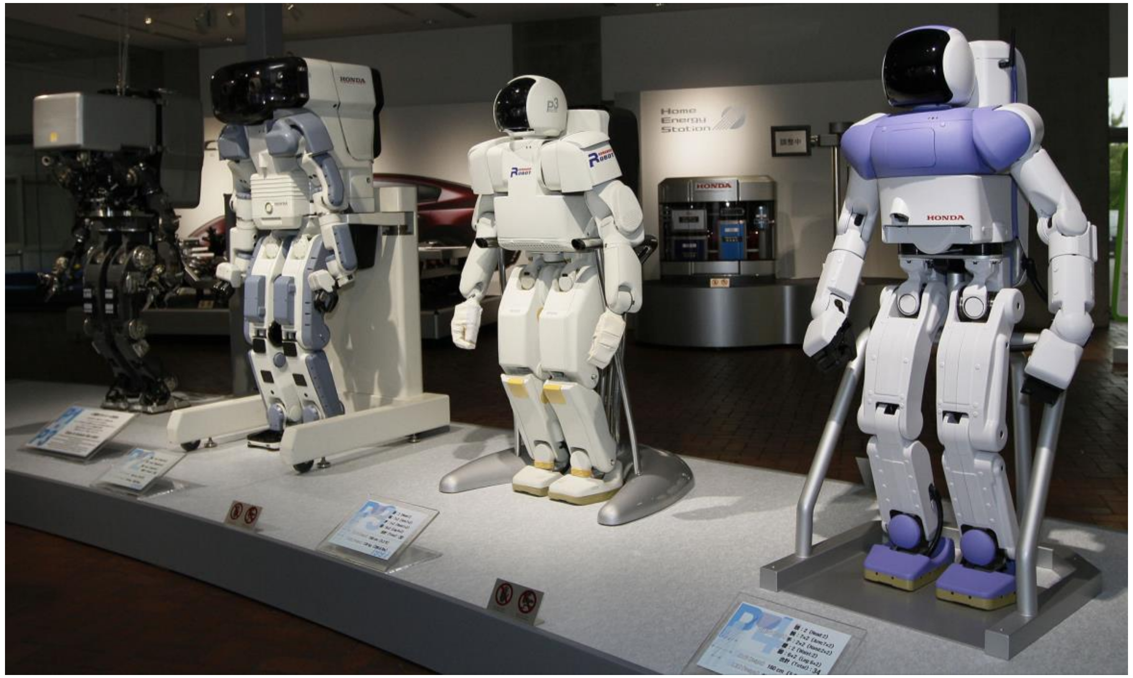
- A goal-oriented machine that can **sense**, **plan** and **act**

  - A robot senses its environment and uses that information, together with a goal, to plan some action

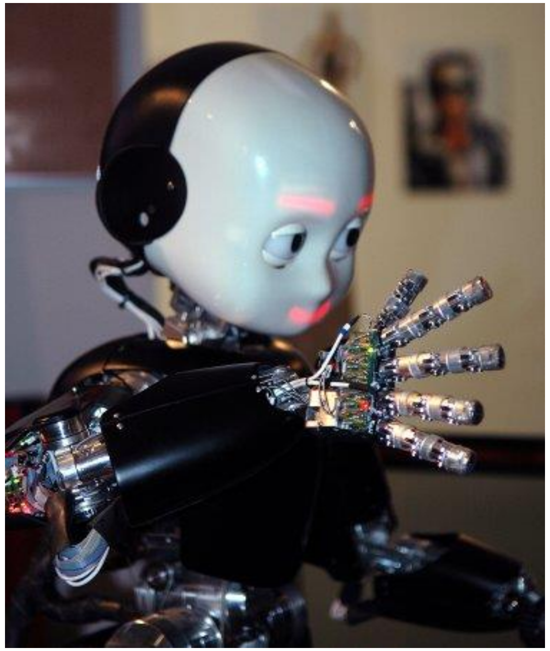
  - The action might be to move the tool of an arm-robot to grasp an object, or it might be to drive a mobile robot to some place

Humanoid Robots

- A humanoid robot is a robot with its body shape built to resemble the human body



Honda P series



iCub robot


HONDA

Robot Manipulators

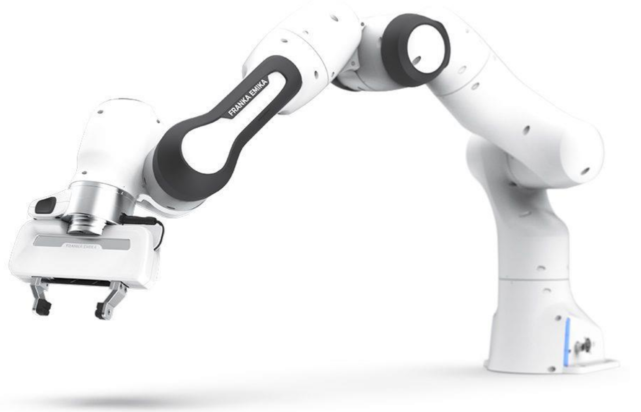
- A device used to manipulate materials without direct physical contact of the operator

ABB

KUKA

FUNUC

Yaskawa



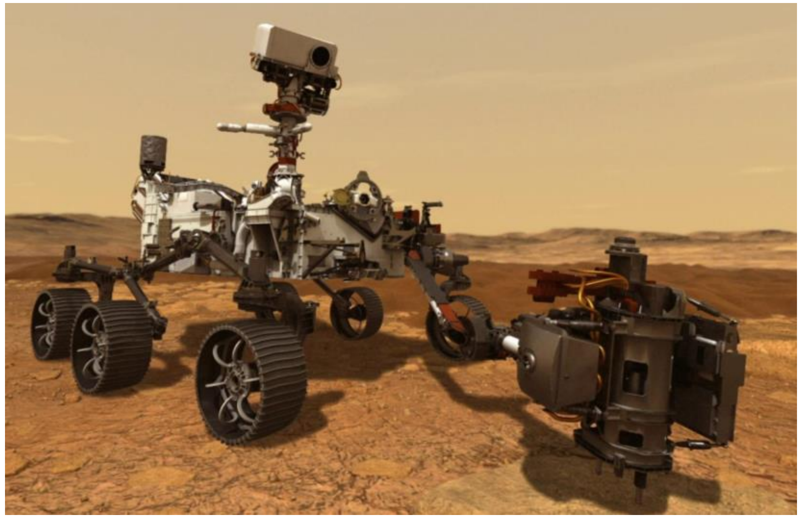
Franka Emika

Wheeled Robots

- Use wheels for locomotion
  - Self-driving cars

Starship Technologies

Amazon Astro Robot



Perseverance Rover

Walking Robots

- Legged robots, use articulated limbs to provide locomotion

Boston Dynamics


Robot Cassie

Boston Dynamics

Other Robots

- Flying robots
  - Drones
- Swimming robots
  - Underwater gliders
- Snake robots

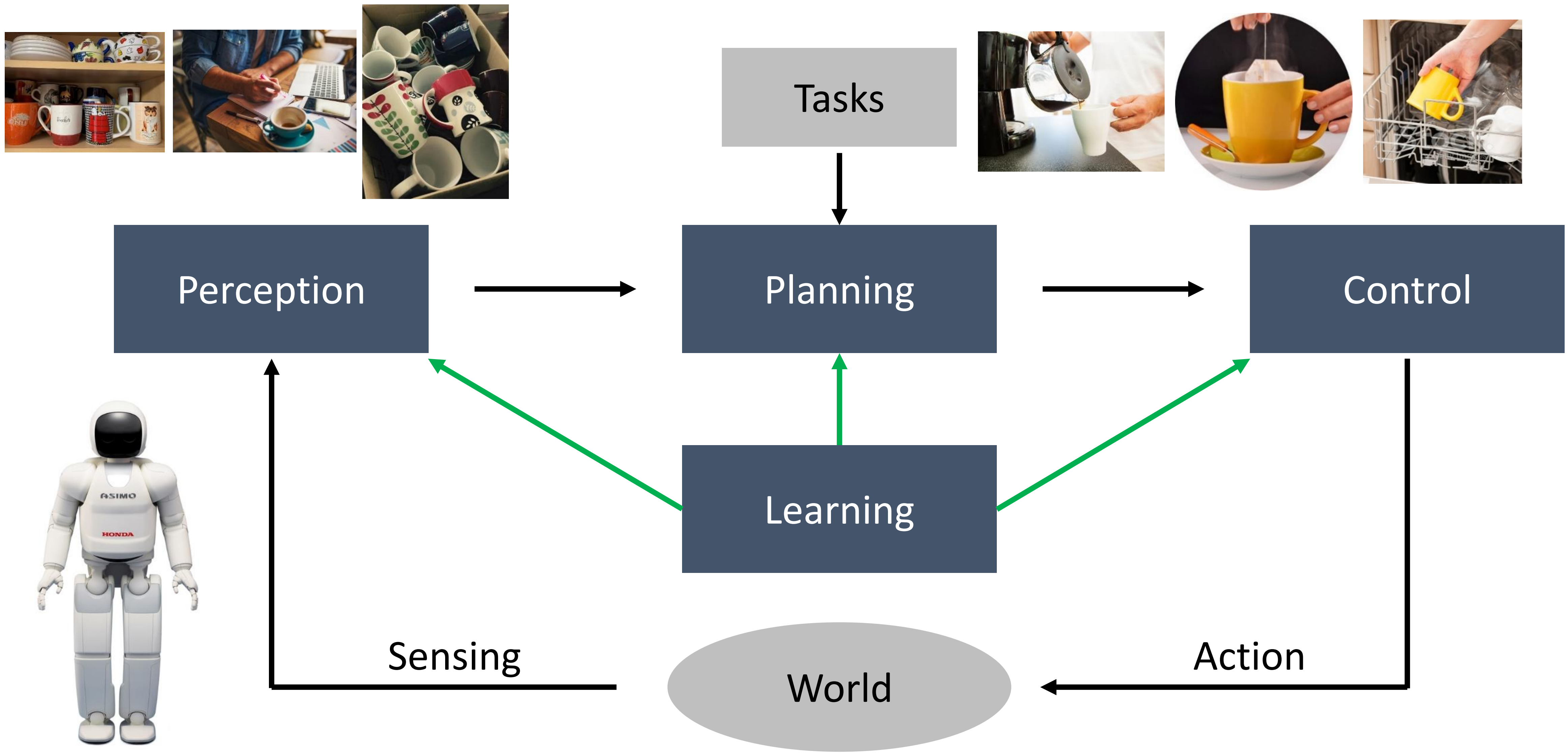
Robotic Fish: *iSplash*-II

Two robot snakes. Left one has 64 motors (with 2 degrees of freedom per segment), the right one 10.

Robots vs. Humans

- Sensing
  - Robots: cameras, IMUs, joint encoders
  - Humans: vision, vestibular, proprioceptive senses
- Control
  - Robots: motors
  - Humans: muscles
- Computation
  - Robots: robot brain, AI?
  - Humans: human brain

Robotic Systems



Computer Vision and Robotics

- Vision provides sensing capability to robots

Vision → Robotics

- Robots can interact with the world to facilitate vision

Robotics → Vision

Robot Manipulation

- The ways robots interact with objects

- Examples
  - Grasping an object
  - Placing an object
  - Pushing an object
  - Opening a door
  - Folding laundry
  - Etc.

https://am.is.mpg.de/research_projects/autonomous-robotic-manipulation

Robot Manipulation

Perception → Planning → Control

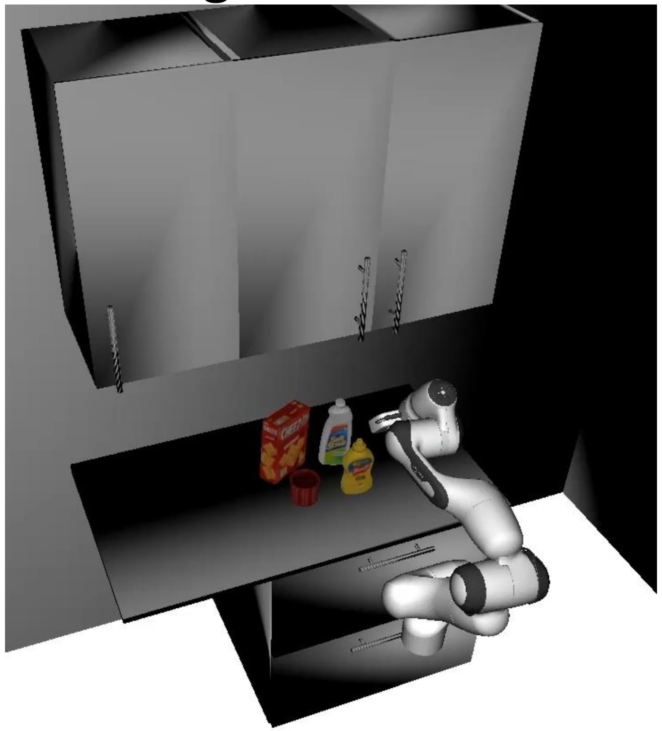
Robust and Accurate

High degree of freedom
Multi-modal grasping

Contact with objects

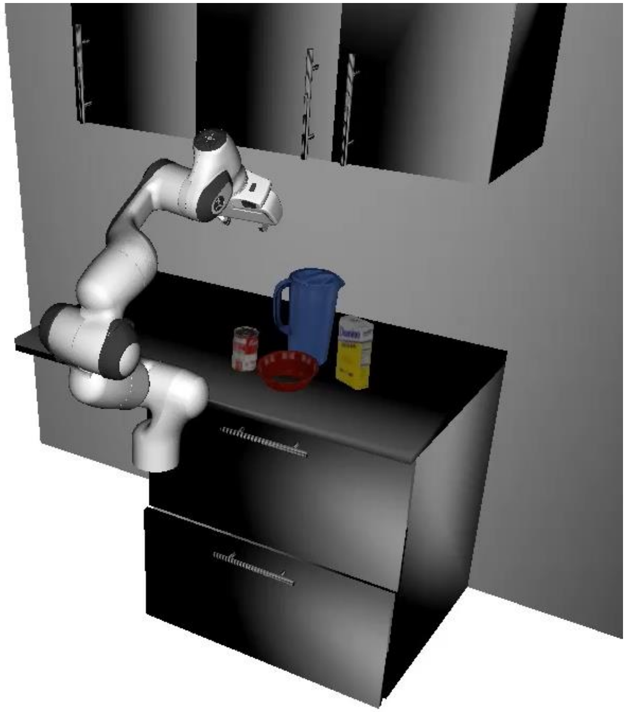
Sensed image

Planning scene



Real world execution

6D Object Pose Estimation for Robot Manipulation

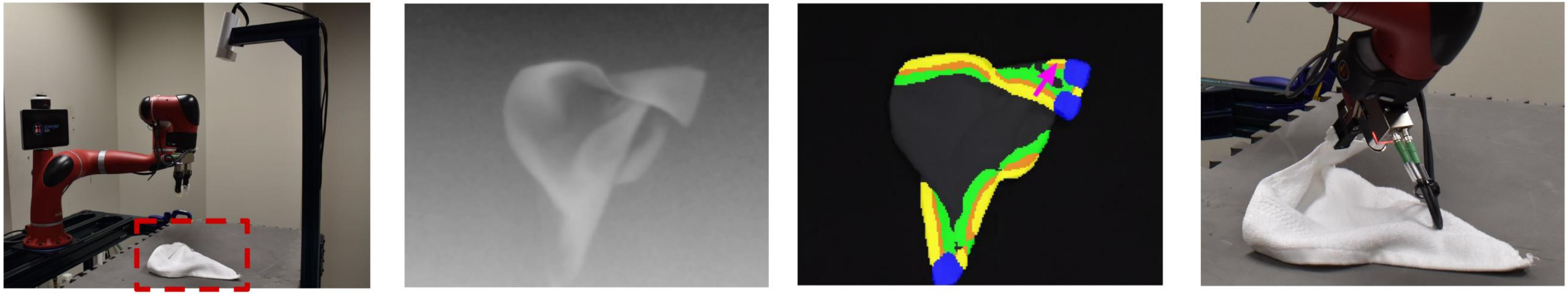


Object Segmentation for Robot Manipulation

Unseen Object Instance Segmentation:
Xie-Xiang-Mousavian-Fox, CoRL’19, T-RO’21
Xiang-Xie-Mousavian-Fox, CoRL’20

6-DOF GraspNet:
Mousavian-Eppner-Fox, ICCV’19

Cloth Segmentation for Robot Manipulation



(a) Initial Setup

(b) Input Depth Image

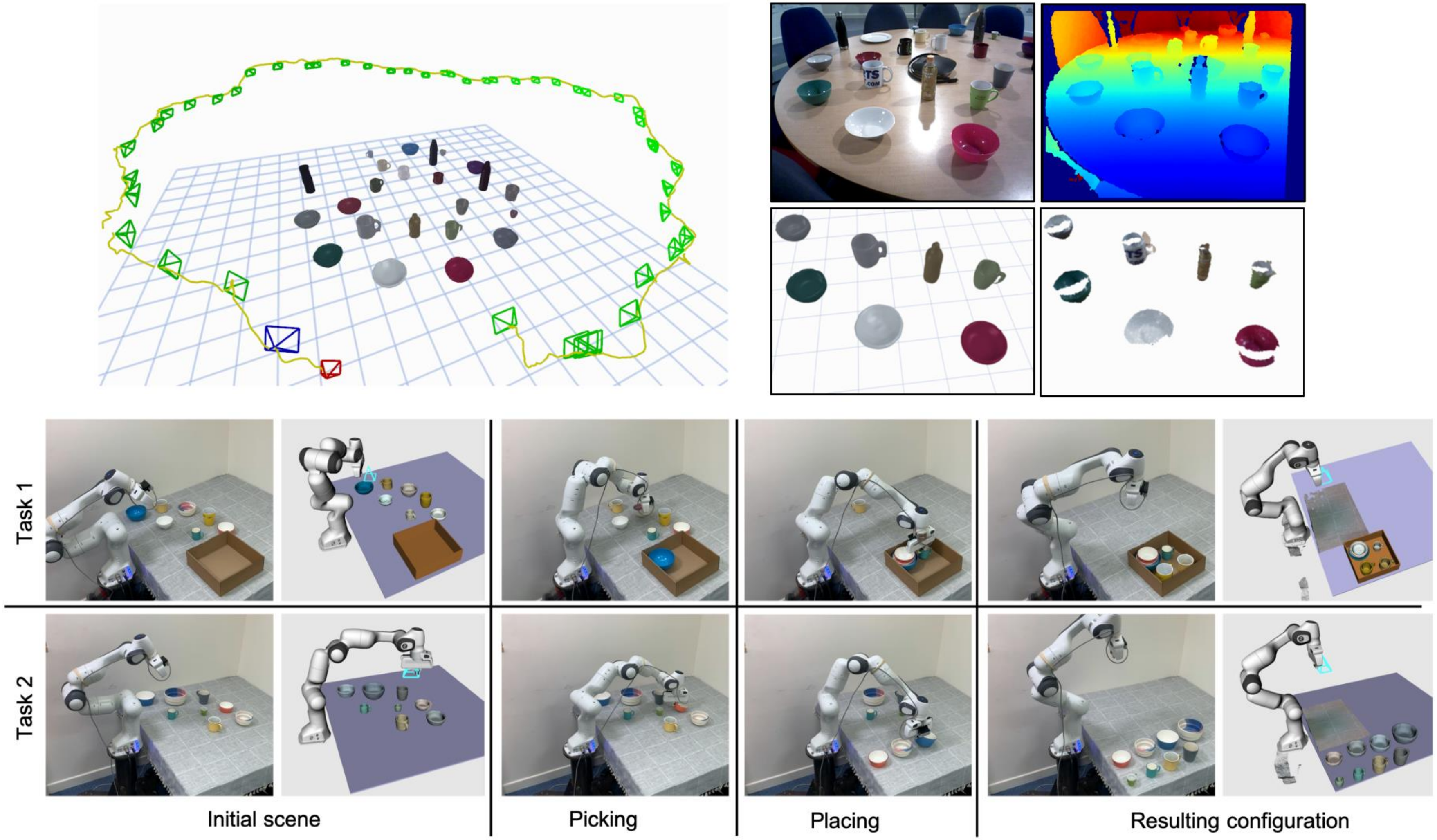
(c) Cloth Segmentation and Grasp Selection

(d) Execute Sliding Grasp

Depth image is segmented into outer edges (yellow), inner edges (green) and corners (blue) using our cloth region segmentation network. Ambiguous regions are colored in orange. Our method selects a grasp location and direction, shown as a magenta arrow.

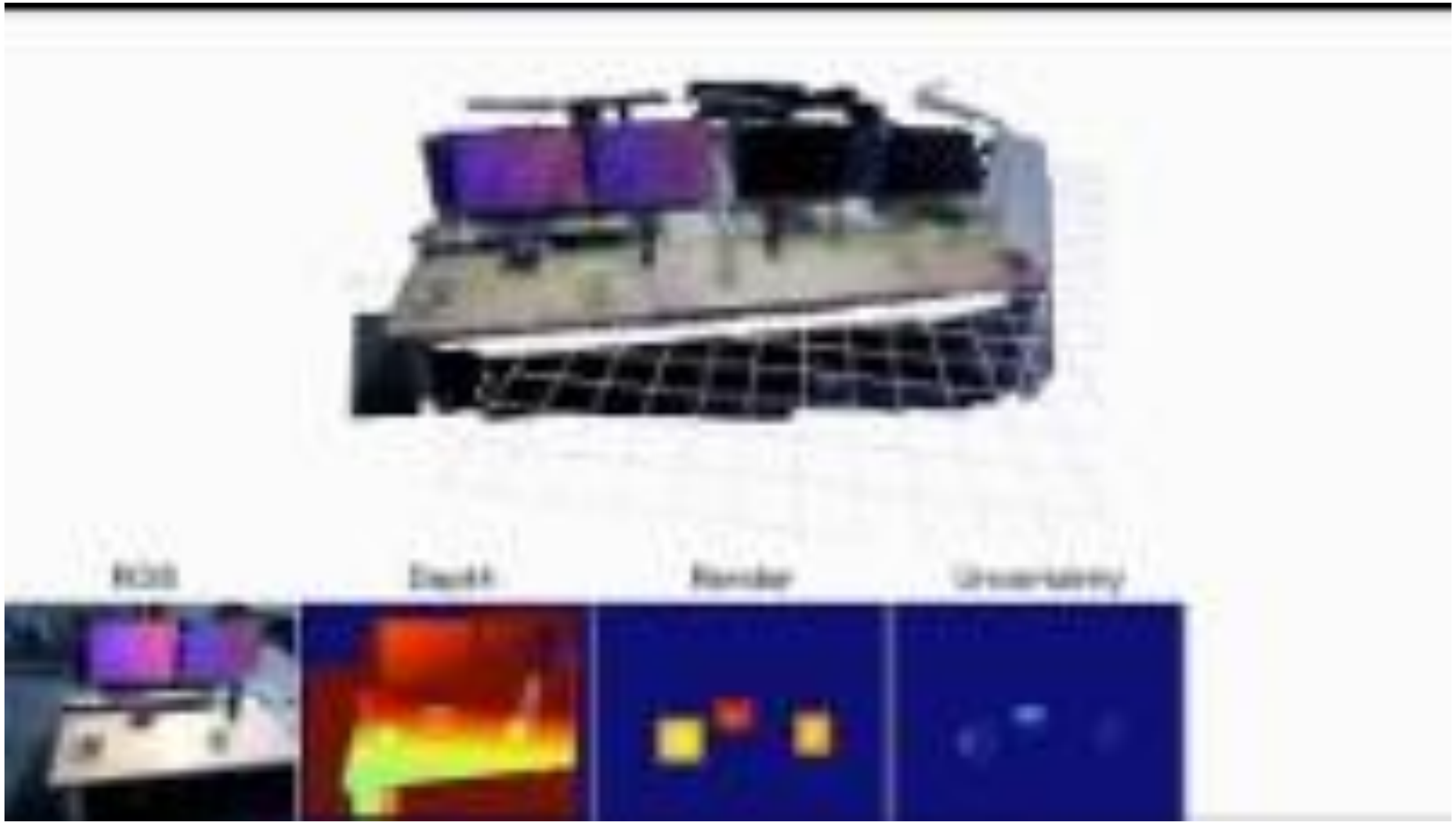
Cloth Region Segmentation for Robust Grasp Selection. Qian et al., IROS, 2020

SLAM for Robot Manipulation



NodeSLAM: Neural Object Descriptors for Multi-View Shape Reconstruction. Sucar et al., 3DV, 2020

SLAM for Robot Manipulation



NodeSLAM: Neural Object Descriptors for Multi-View Shape Reconstruction. Sucar et al., 3DV, 2020

Robot Navigation

- Go from A to B without hitting anything

Perception → Planning → Control

Simultaneous localization and mapping (SLAM)

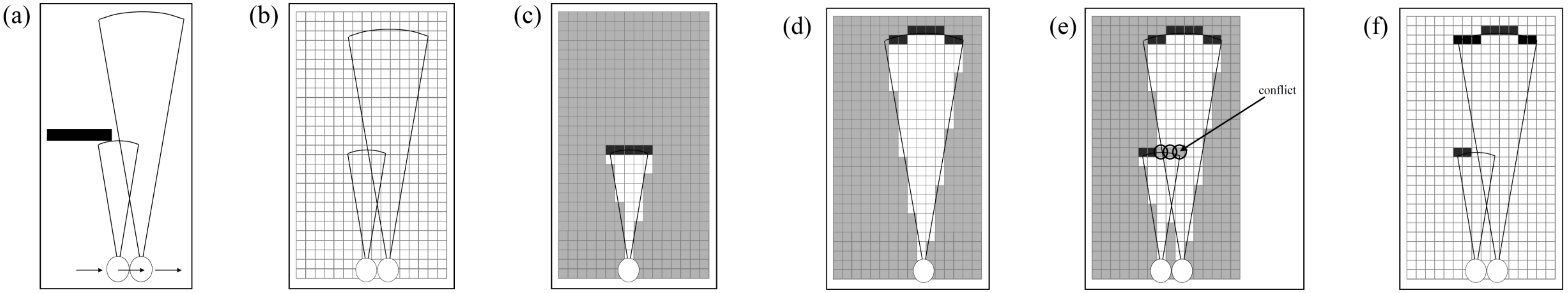
Path planning

Path following

Laser-based SLAM
2D occupancy grid map

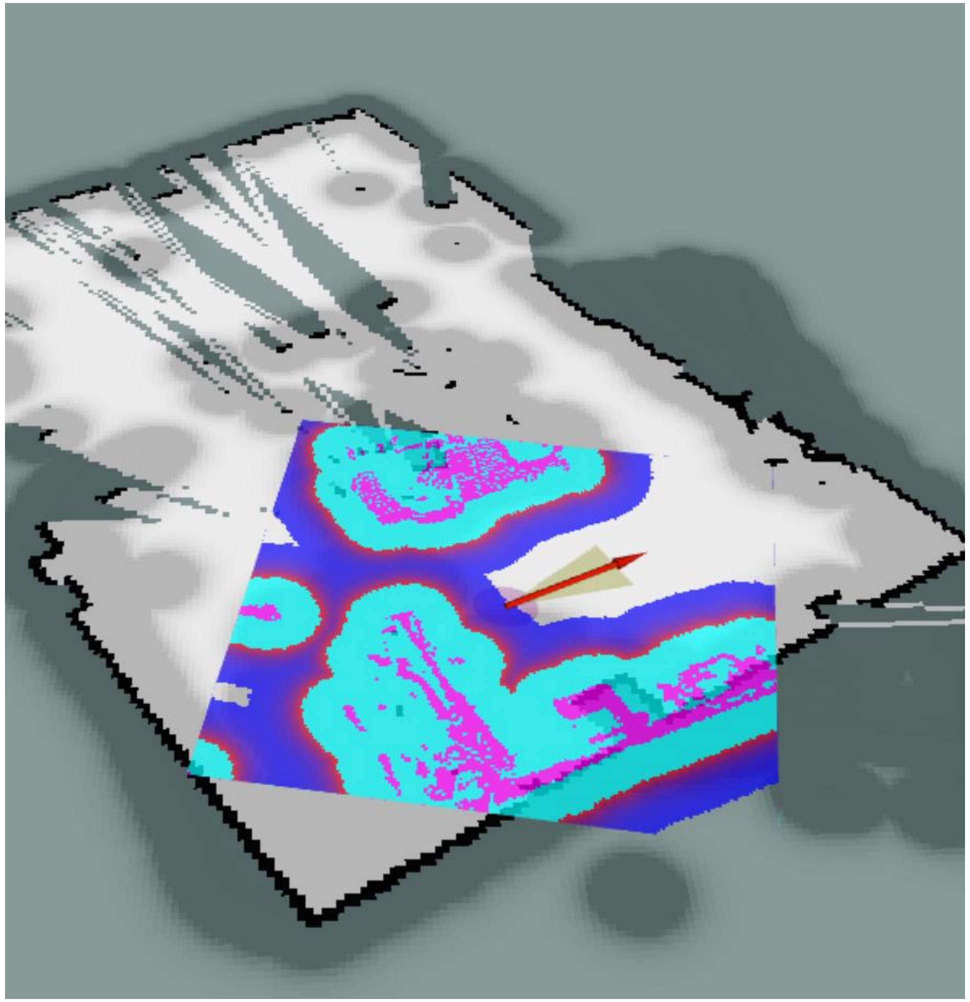
Occupancy Grid Mapping

- Occupancy grid
  - Status: unknown, occupied, empty



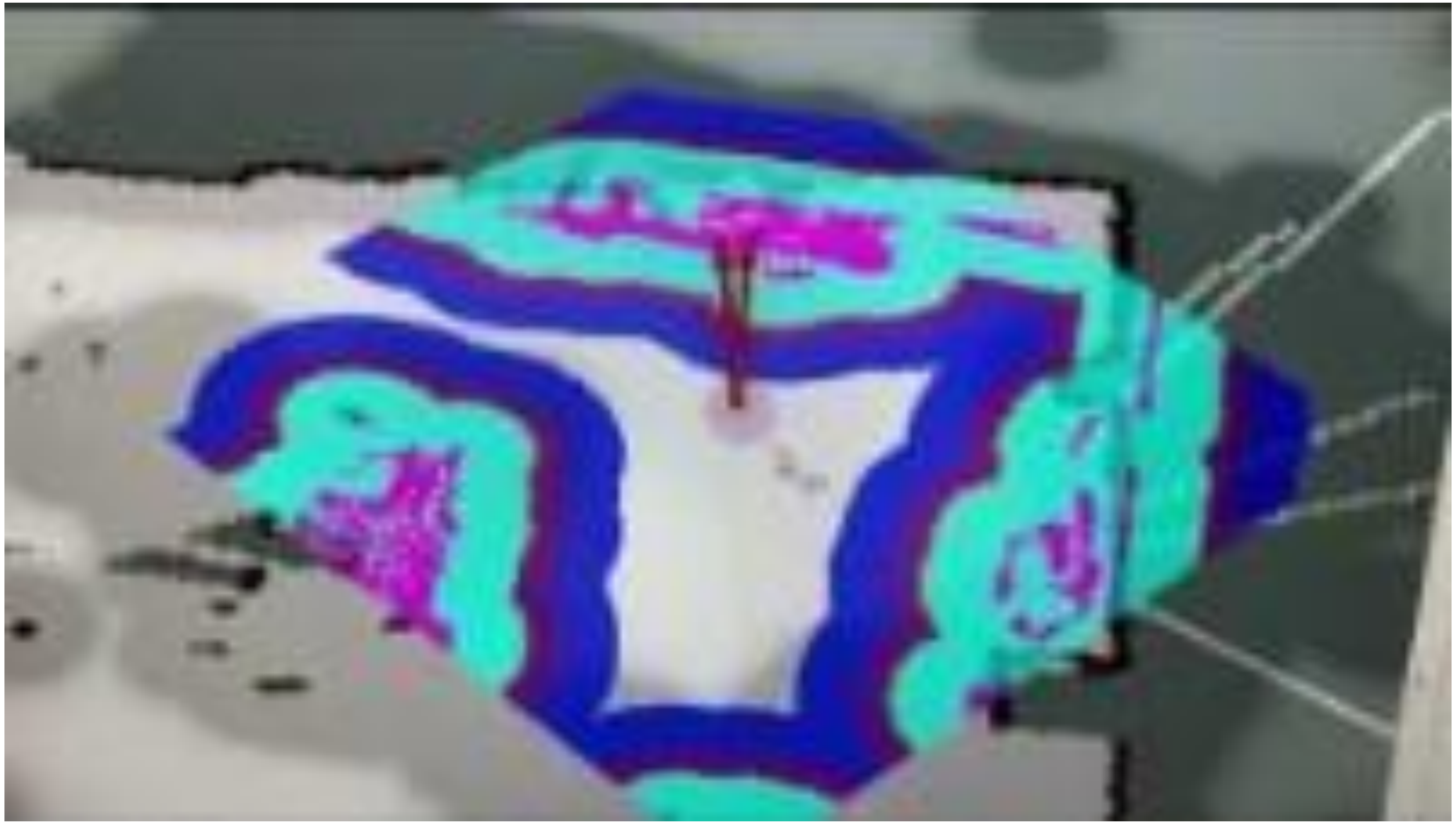
Learning Occupancy Grid Maps With Forward Sensor Models. Sebastian Thrun, 2002

Occupancy Grid Mapping



Navigation Demo using ROS

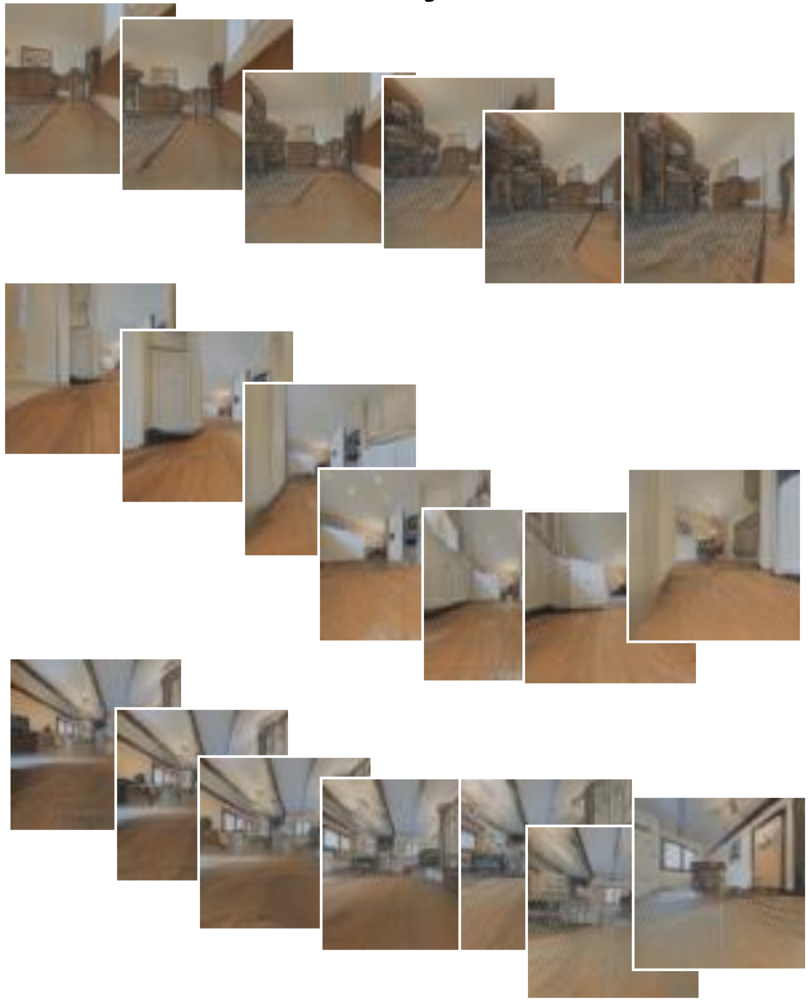
Credit: Gagan Bhat



Topological Navigation

Meng-Ratliff-Xiang-Fox, ICRA’19, ’20
Meng-Xiang-Fox, RA-L’21

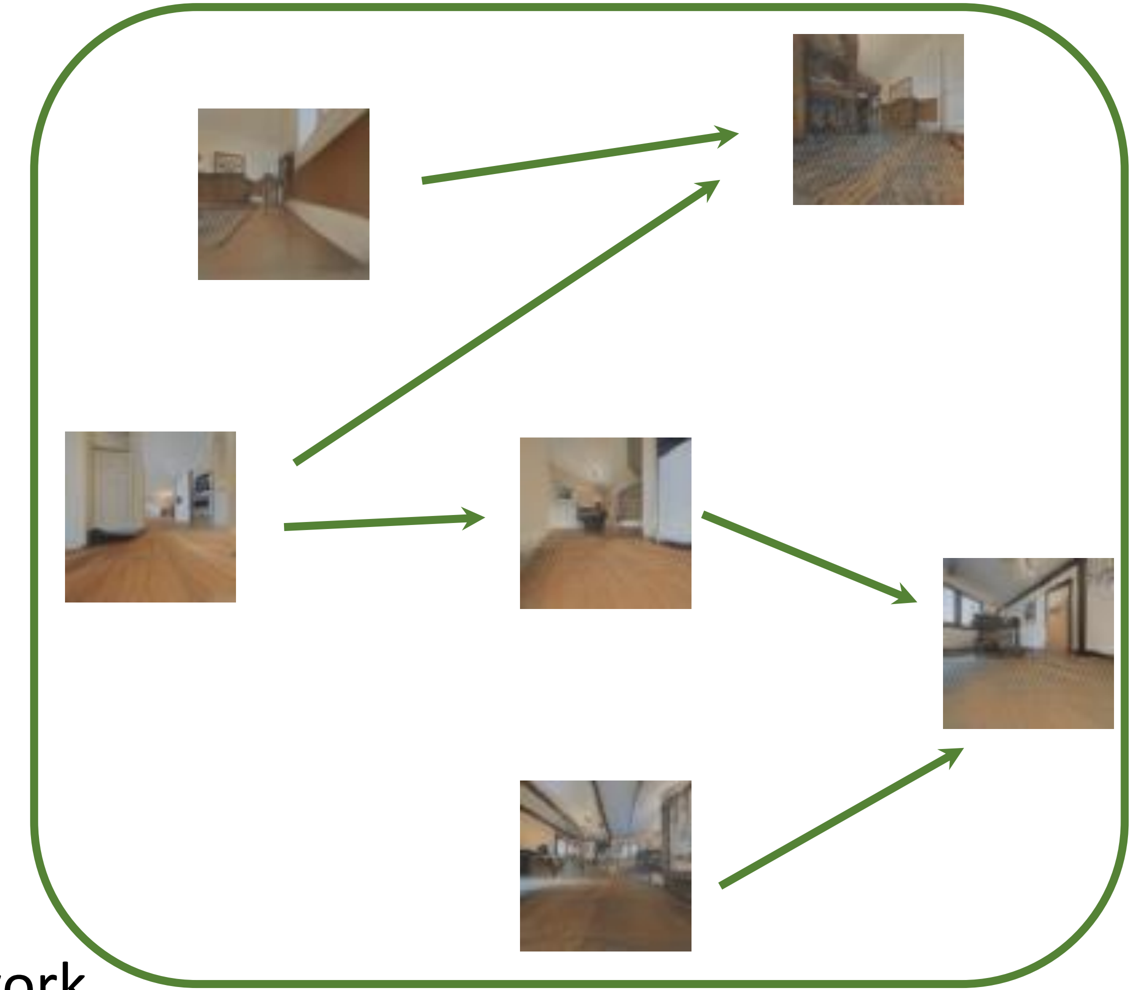
Dense Trajectories



Reachability Estimator

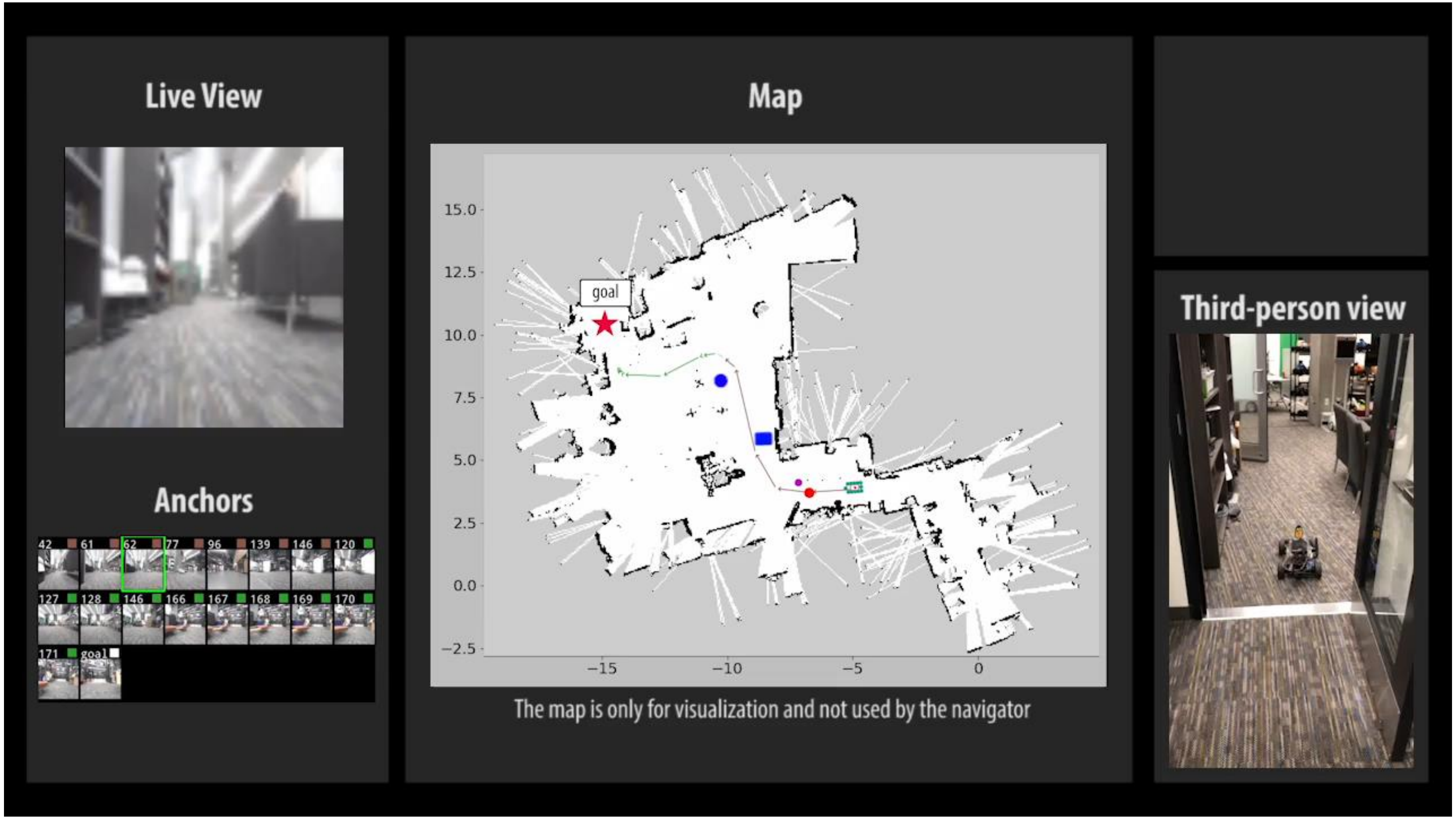
Local Controller
A learned neural network

Sparse Topological Map



Robot Navigation

Meng-Ratliff-Xiang-Fox, ICRA'19, '20
Meng-Xiang-Fox, RA-L'21



Robot Teleoperation

https://www.shadowrobot.com/jeff-bezos-tries-our-tech/

Hand Pose Estimation for Robot Teleoperation

DexPilot: Vision Based Teleoperation of Dexterous Robotic Hand-Arm System. Handa et al. ICRA’21.

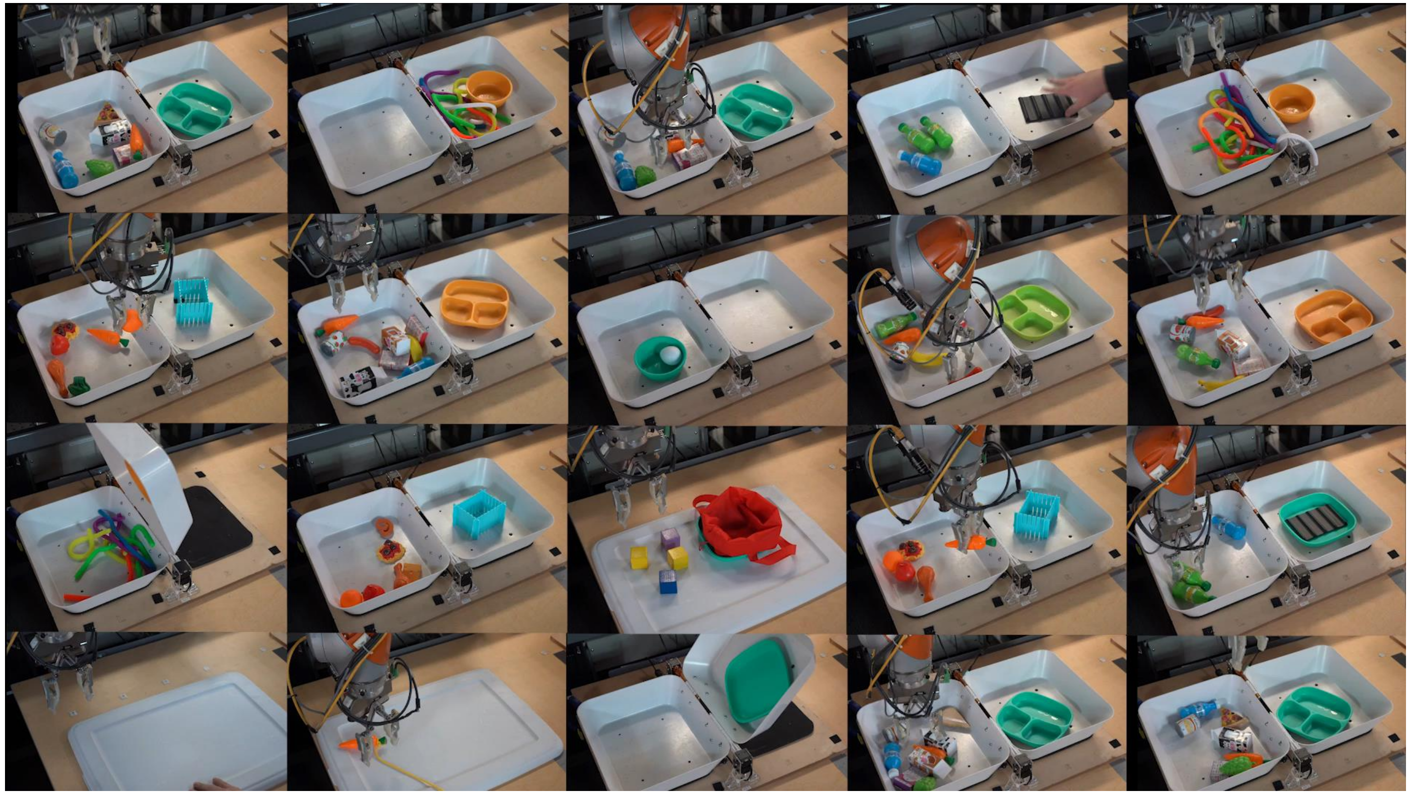
Robot Learning

- How can robots learn various skills?
  - Navigation
  - Manipulation

- Reinforcement learning
  - Learning from trial and error

- Imitation learning
  - Learning from demonstrations

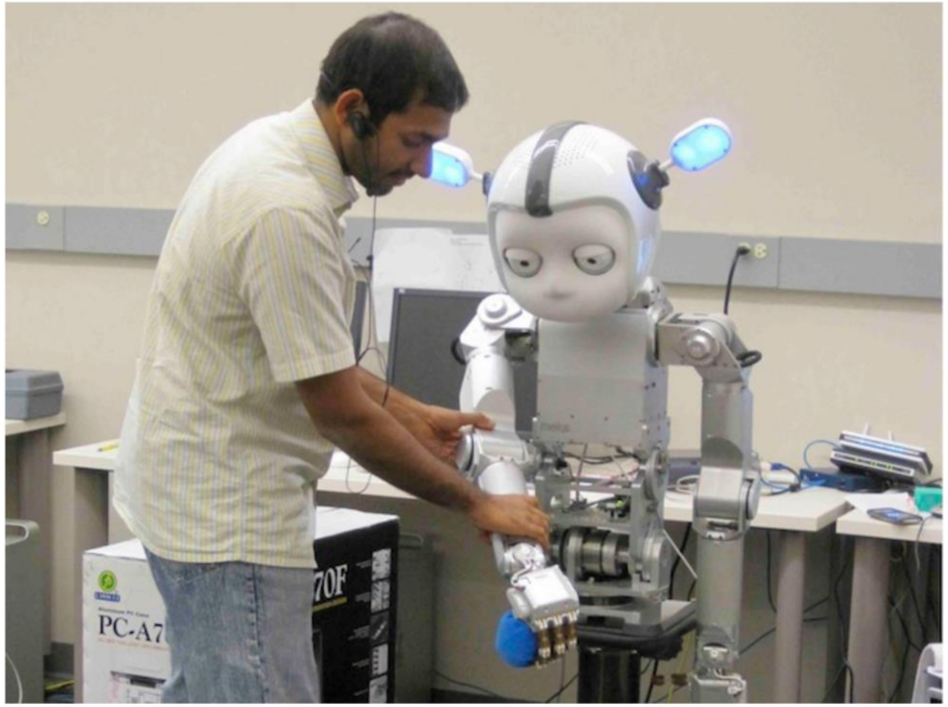
Reinforcement Learning



https://ai.googleblog.com/2021/04/multi-task-robotic-reinforcement.html

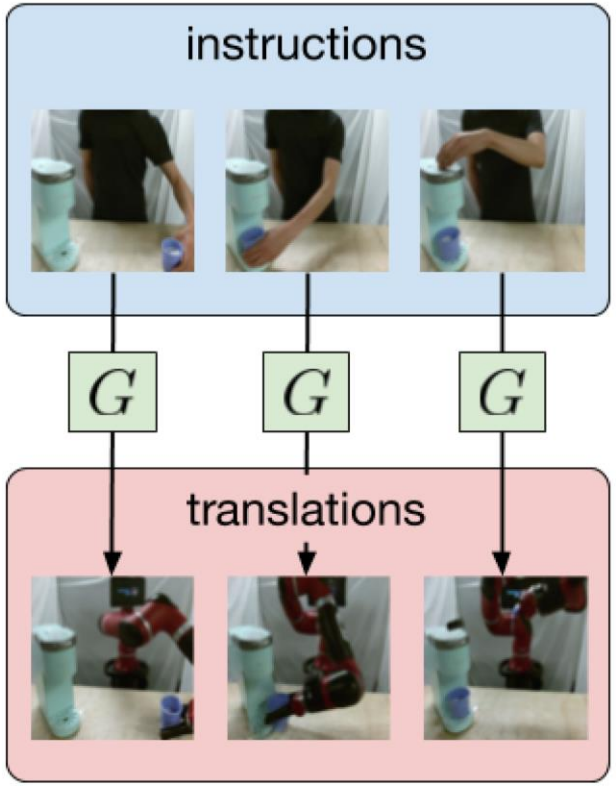
Kinesthetic Teaching

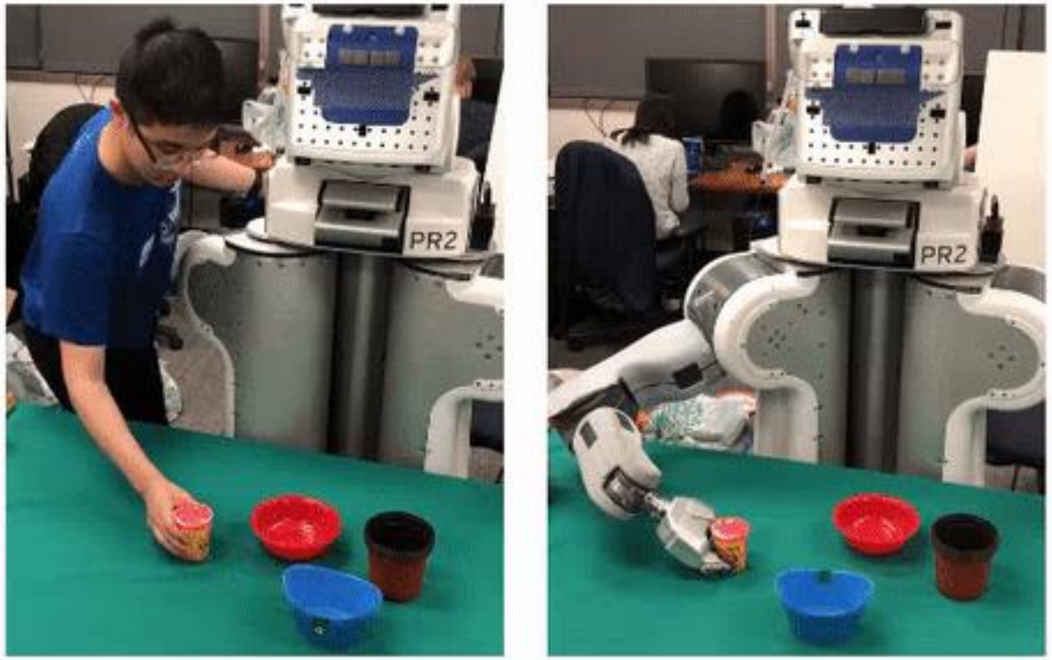
- A human teacher physically guides the robot in performing the skill



Keyframe-based Learning from Demonstration. Akgun et al., International Journal of Social Robotics , 2011.

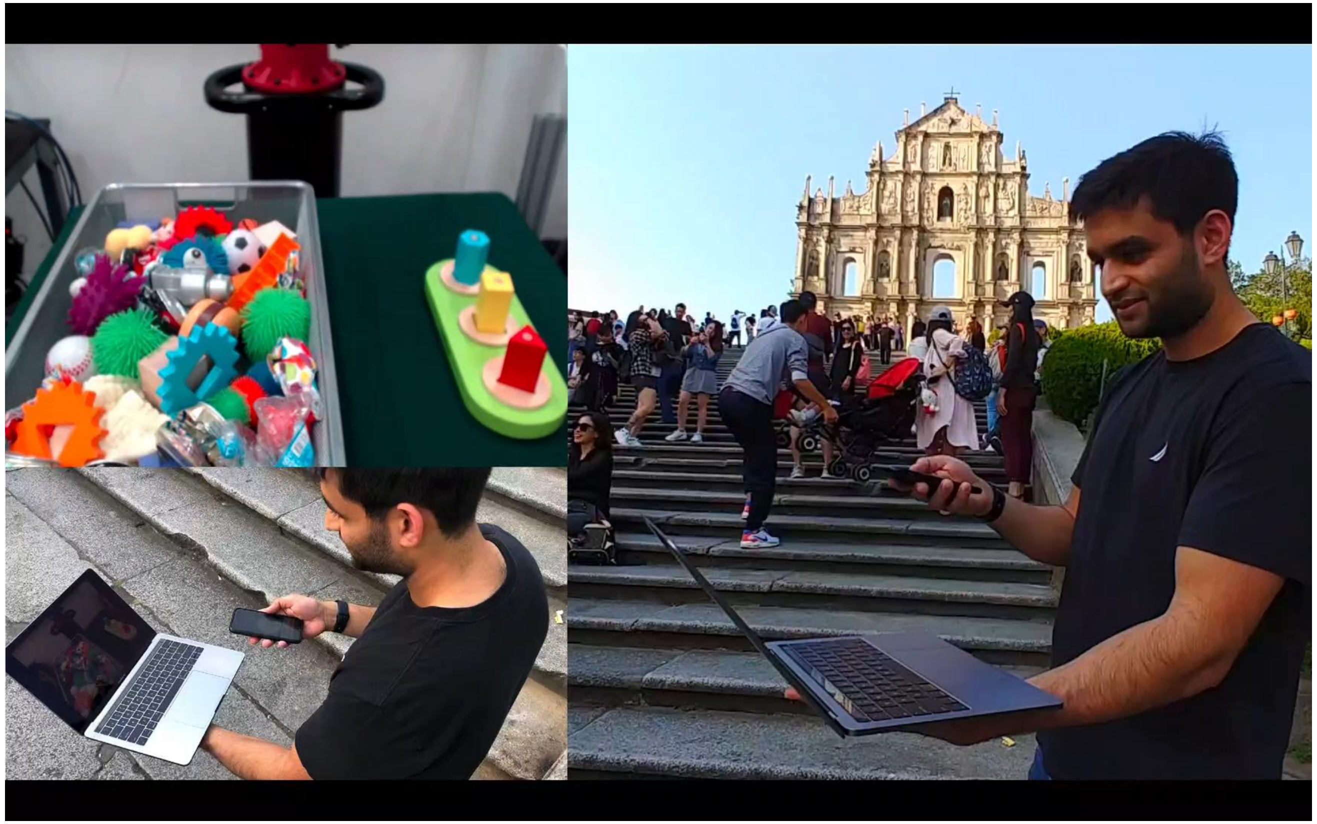
Watching Human Demonstrations





AVID: Learning Multi-Stage Tasks via Pixel-Level Translation of Human Videos. Smith et al., arXiv'20.

Demonstrations from Teleoperation



RotoTurk: https://roboturk.stanford.edu/realrobotdataset#dataset

Demonstrations from VR

- Use VR in teleoperation to generate demonstrations

Deep Imitation Learning for Complex Manipulation Tasks from Virtual Reality Teleoperation. Zhang et al., in arXiv'18.

Demonstrations from VR

https://techxplore.com/news/2017-11-startup-robots-puppets.html

TRI Robotics Example

TOYOTA
RESEARCH INSTITUTE

Summary

- Vision is important for robots to sense the world
- Robots can interact with the world to facilitate vision
  - Interactive perception
- Core problems in robotics
  - Manipulation
  - Navigation
  - Skill learning

Further Reading

- 6-DOF GraspNet: Variational Grasp Generation for Object Manipulation, 2019 https://arxiv.org/abs/1905.10520
- Cloth Region Segmentation for Robust Grasp Selection, 2020 https://arxiv.org/abs/2008.05626
- Learning Occupancy Grid Maps With Forward Sensor Models, 2002 https://www.cs.cmu.edu/~thrun/papers/thrun.occ-journal.pdf
- Scaling Local Control to Large Scale Topological Navigation, 2020 https://arxiv.org/abs/1909.12329
- DexPilot: Vision Based Teleoperation of Dexterous Robotic Hand-Arm System, 2019 https://arxiv.org/abs/1910.03135
- AVID: Learning Multi-Stage Tasks via Pixel-Level Translation of Human Videos, 2019 https://arxiv.org/abs/1912.04443
- Deep Imitation Learning for Complex Manipulation Tasks from Virtual Reality Teleoperation, 2017 https://arxiv.org/abs/1710.04615