GroceryBee

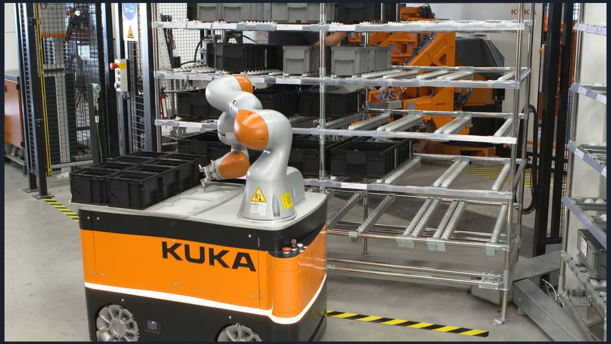
Group T2323
Members:
Efe Beydoğan 21901548
Mert Barkın Er 21901645
Emir Melih Erdem 21903100
Arda Önal 21903350
Eren Polat 21902615

Background Information

- Surge in the number of online grocery orders
- Turkey’s online supermarket sales grew by 434% in 2020
- US: 70% of shoppers state they would buy groceries online by 2022
- GroceryBee: automate & expedite online shopping process
- Automate collection of items from shelves

Background Information

- Improve current developments: sustained innovation
- Transformation of system (humans -> robots) to optimize gathering items
- Emphasize process type of innovation: aim to enhance inner processes of markets
- Assume an existing system receives and relays the order info to robot
- Delivery will be done by store employees
- Our project: focus on traversing store, collect items

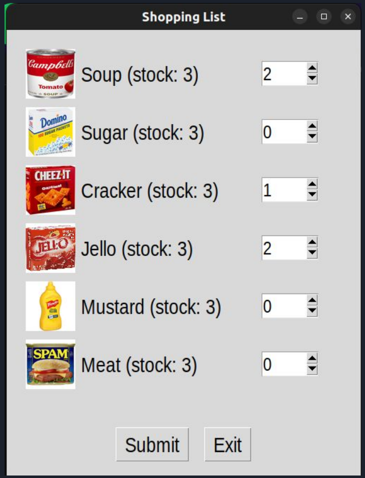


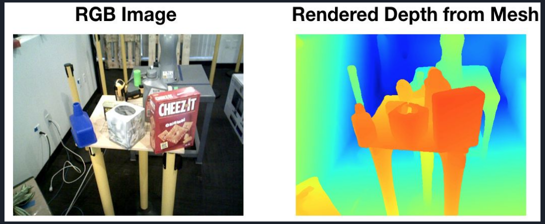
Project Plan

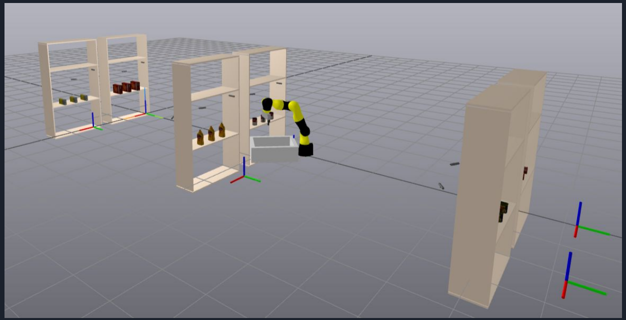
- Collect items for online orders
- Navigate the supermarket without crashing
- Detect object, pick object up with gripper, bring object back for delivery
- Calculate optimal path to collect orders
- 3 technical pillars of our project:
    - Motion planning
    - Task planning
    - Perception

Current System

- Autonomous agent
- Make sequential decisions in an uncertain environment
- Not possible to plan everything beforehand
- Technical features:
  - **Motion planning:** DiffIK, Trajectory Optimization, Mobile base
  - **Task planning:** Travelling salesman problem
  - **Perception:** deep learning—used to segment images
- Additional features:
  - Dynamic environment generation
  - Shopping list input



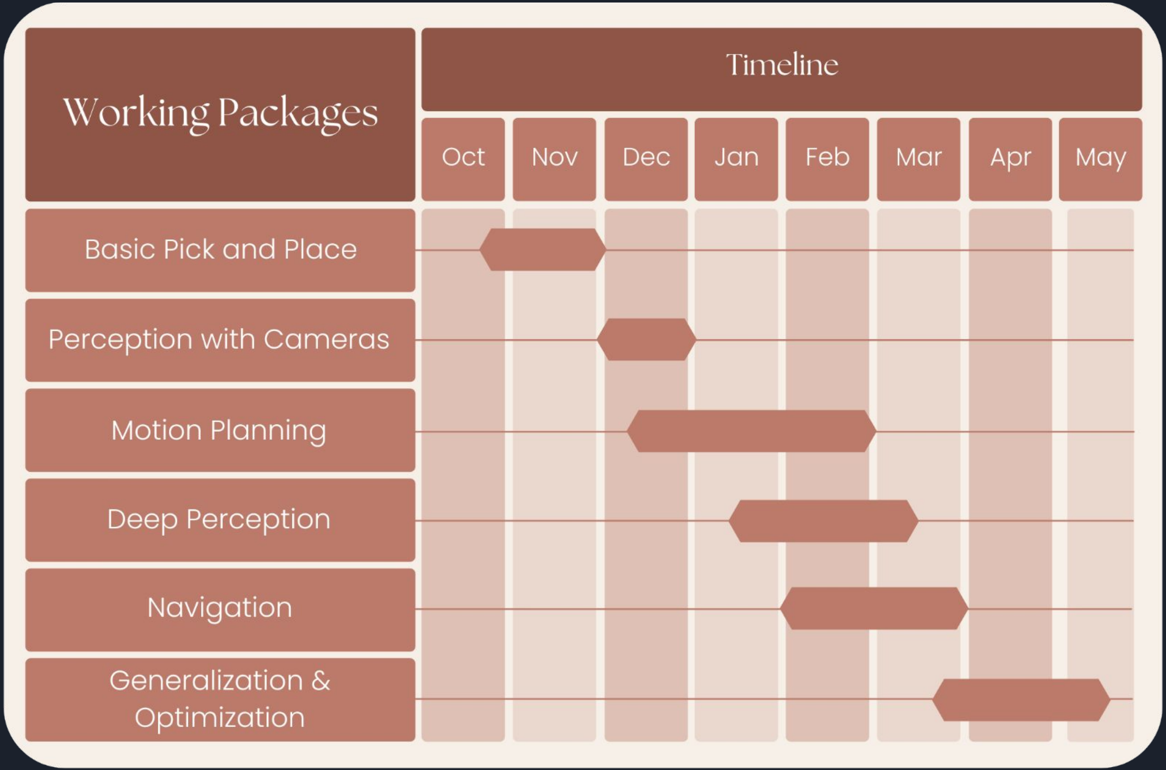




Used Technologies

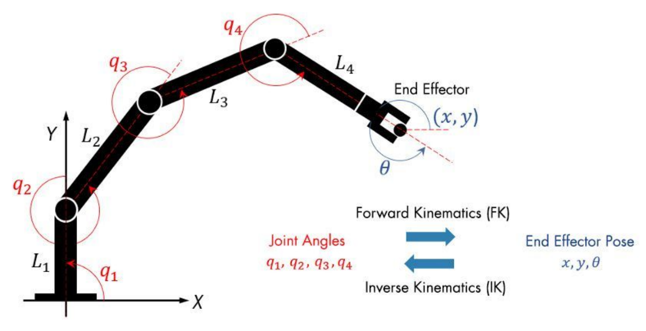
- Drake—C++/Python toolbox for Robotics
- Allows users to create virtual simulations
- Includes a range of algorithms and functions
- Has tools for sensing and perception
- Extensive math packages
- Physics engine
- LangSAM and BERT models for segmentation

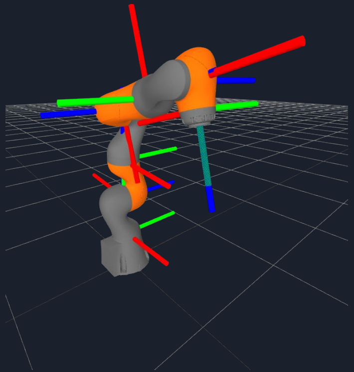
Gantt Chart


Working Packages
Timeline
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
Basic Pick and Place
Perception with Cameras
Motion Planning
Deep Perception
Navigation
Generalization & Optimization

Motion Planning: DiffIK

- Forward kinematics: calculating gripper pose from joint angles
- Inverse kinematics: calculating the joint angles to achieve a desired gripper pose
- Differential inverse kinematics (DiffIK)

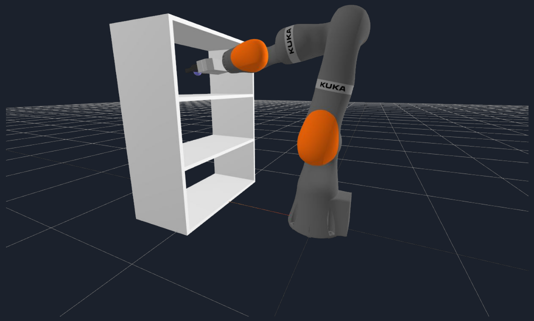




Motion Planning: Kinematic Trajectory Optimization

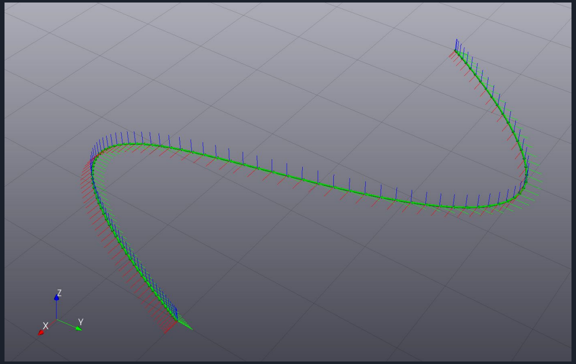
- Solve for all joint angles in the same optimization
- "find a trajectory, q(t), that moves the gripper from the start to the goal in minimum time"
- Local minima

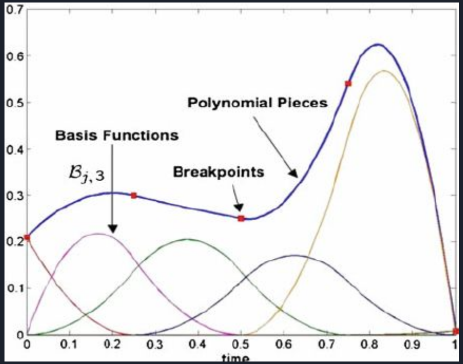
$$\begin{aligned}
\min_{\alpha,T} \quad & T, \\
\text{subject to} \quad & X^{G_{start}} = f_{kin}(q_\alpha(0)), \\
& X^{G_{goal}} = f_{kin}(q_\alpha(T)), \\
& \forall t, \quad |\dot{q}_\alpha(t)| \le v_{max}.
\end{aligned}$$



Motion Planning: Mobile Base

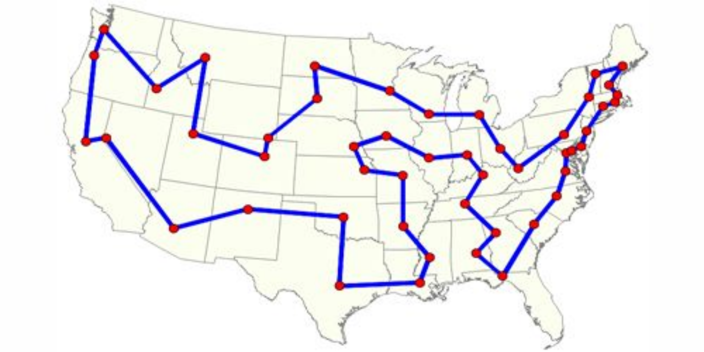
- From specified key positions to trajectories
- Piecewise polynomial trajectories
- Cubic splines for smooth trajectories





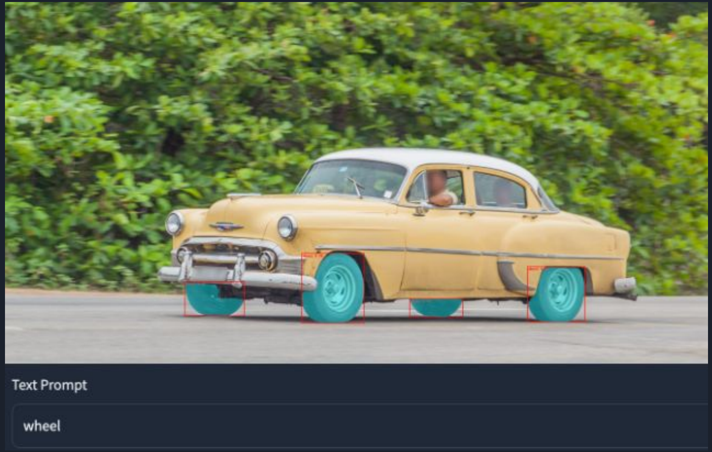
Task Planning

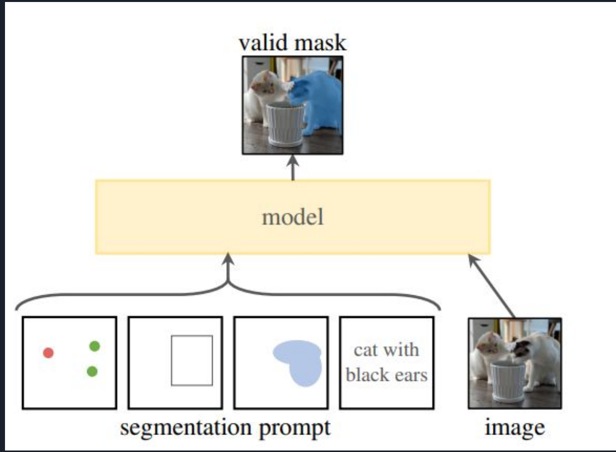
- Several states: waiting, going to shelf, picking, etc.
- In which order should the robot visit the shelves? *Traveling Salesman Problem*
- Visit the desired shelves in such an order that minimizes the route



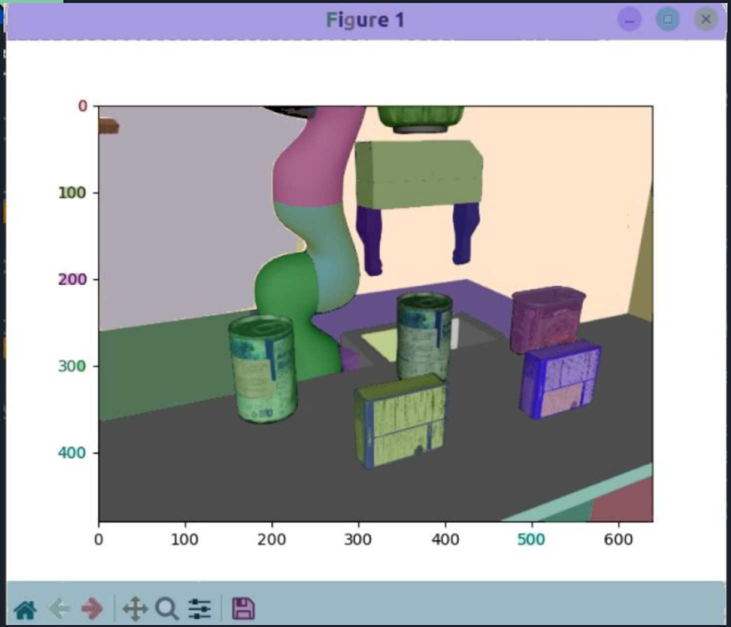
Perception

- Hybrid system (DL and traditional CV algorithms)
- SOTA models to detect desired objects **autonomously**
- SAM (Segment Anything Model) by Meta (2023)
- LangSAM, combining text prompts with SAM model (2023)
- Using both GroundingDINO and SAM, trained using existing bounding box annotations and aims at detecting arbitrary classes with the help of language generalization.

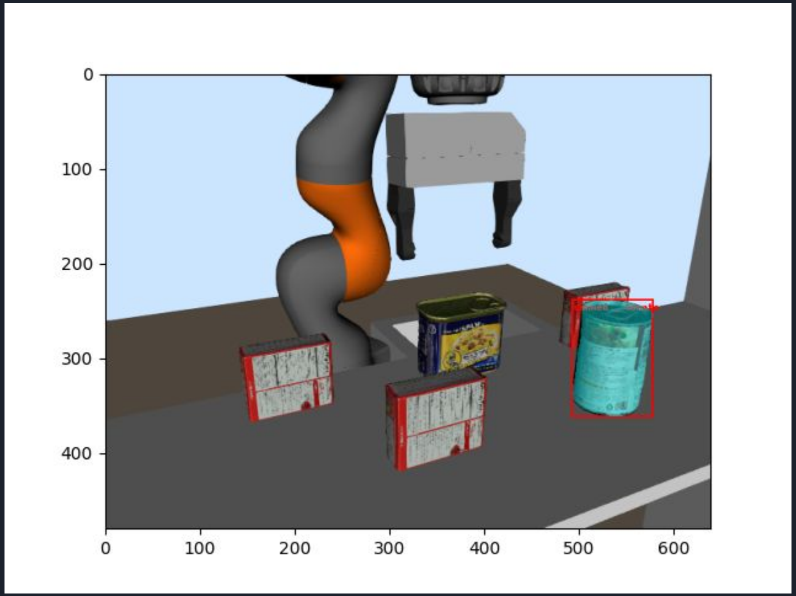




Perception



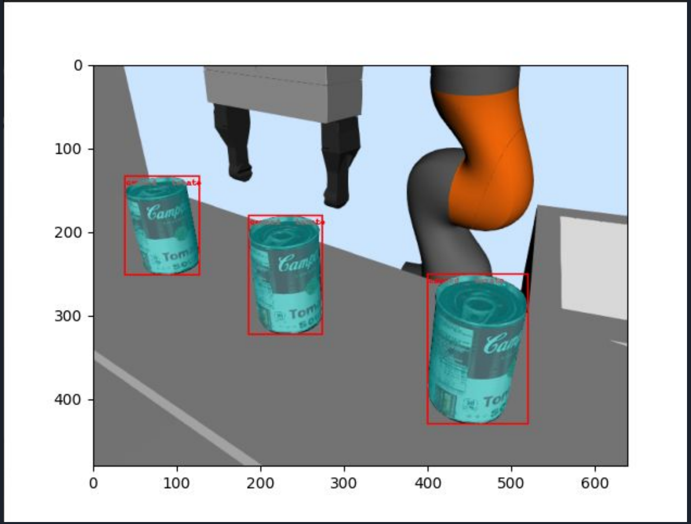
SAM



LangSAM with text prompt

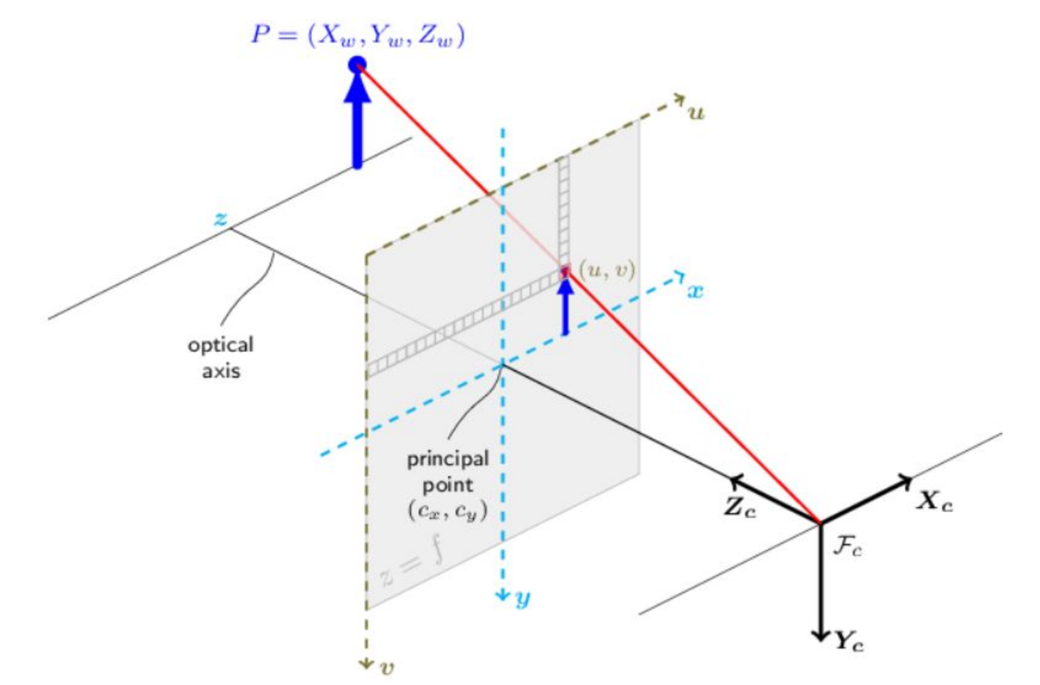
Perception

- Normally, we would crop the shelf by its hard coded coordinates
- Now, we crop the output mask of Lang-SAM, convert 2D to 3D and feed it to our traditional algorithms (ICP...) for grasp selection



Perception

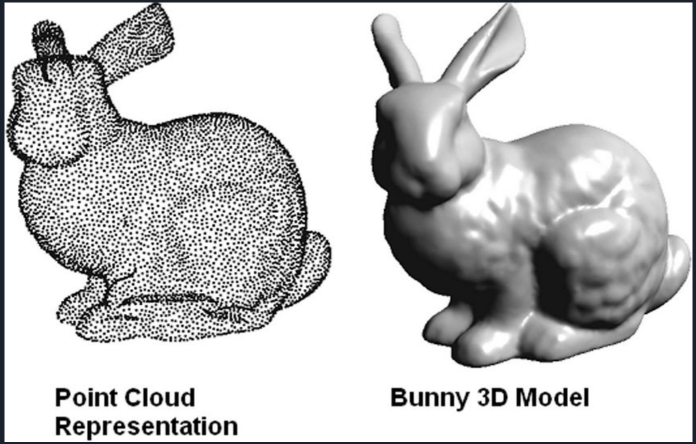
- After objects are segmented, calculate top left and bottom right pixels of segmented mask
- Convert 2D pixel coordinates of the RGB image to 3D relative position using the **pinhole camera model** and **camera intrinsics (optical center, focal length)**
- Convert 3D coordinates relative to the camera to world coordinates

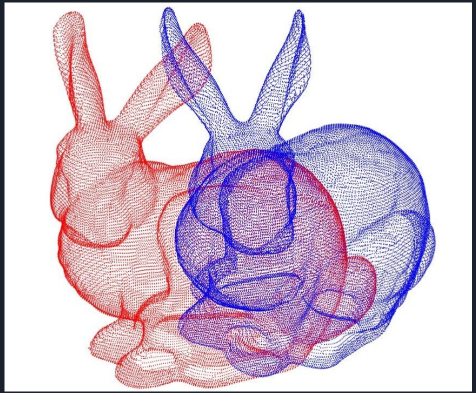


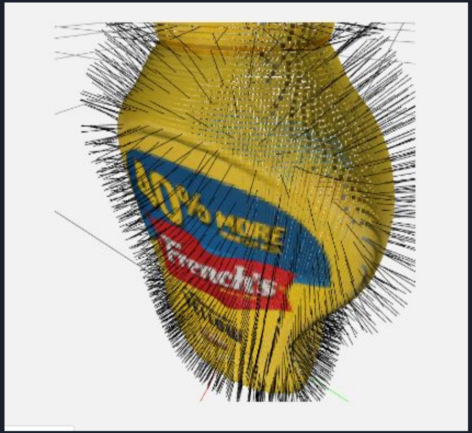
Pinhole camera model

Perception

- Retrieve the **point clouds** belonging to the segmented 3D volume space
- Merge point clouds and use ICP (iterative closest point) to match objects with perceived objects
- Find gripping points
- Overall process is completed autonomously



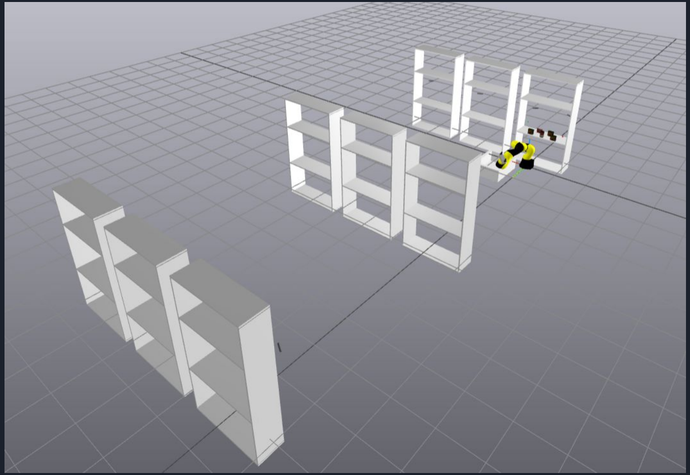


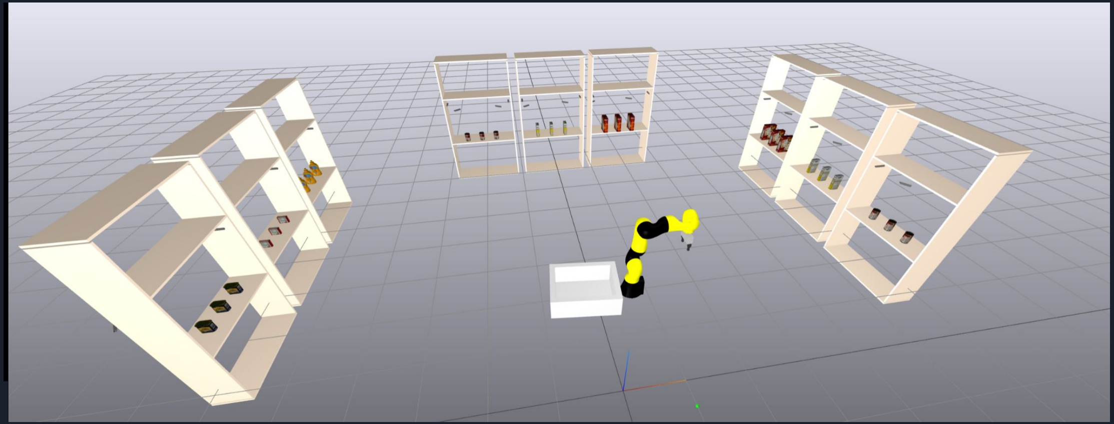


Normal vectors of a sample object for surface curvatures

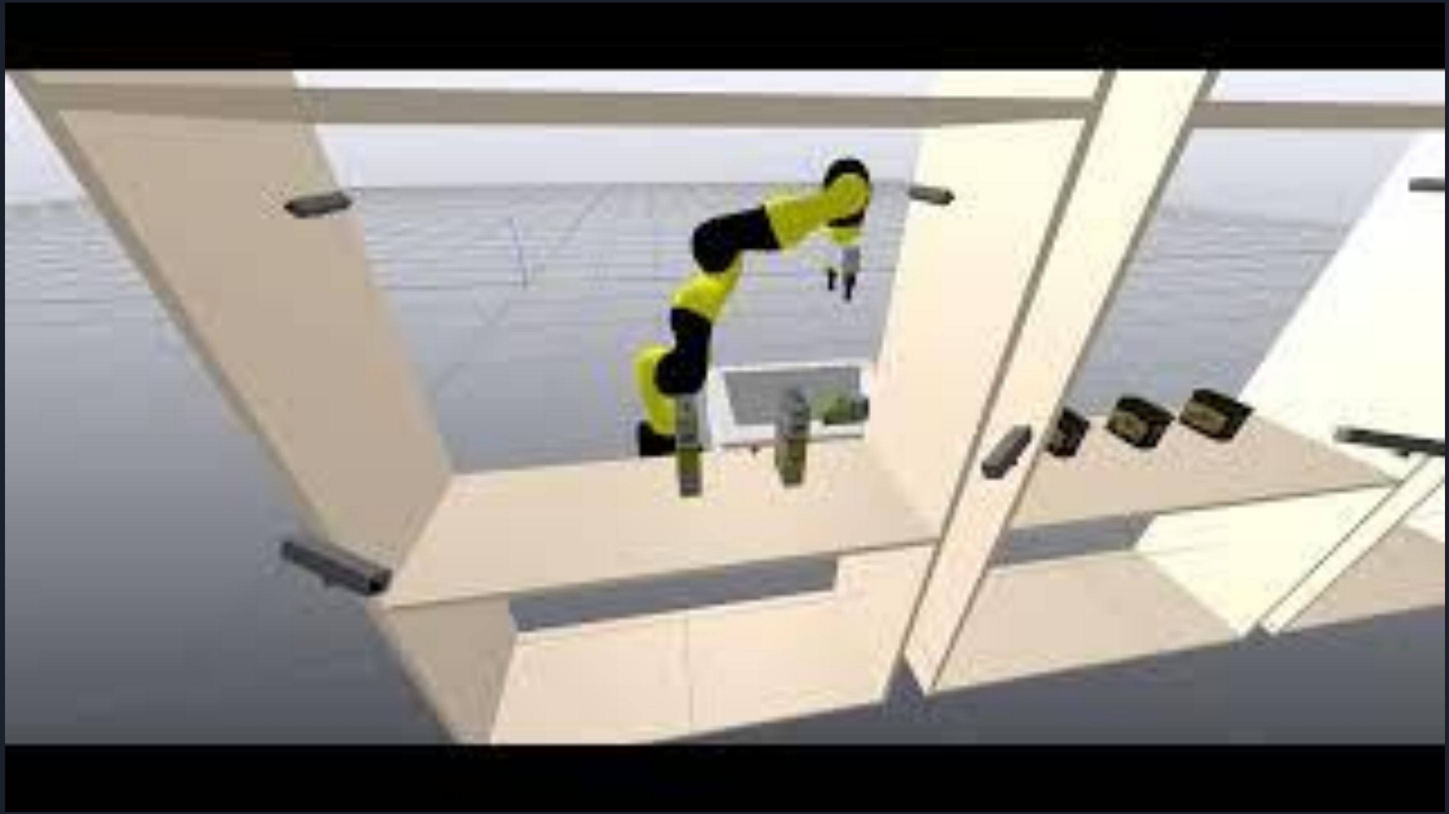
Environment Generation

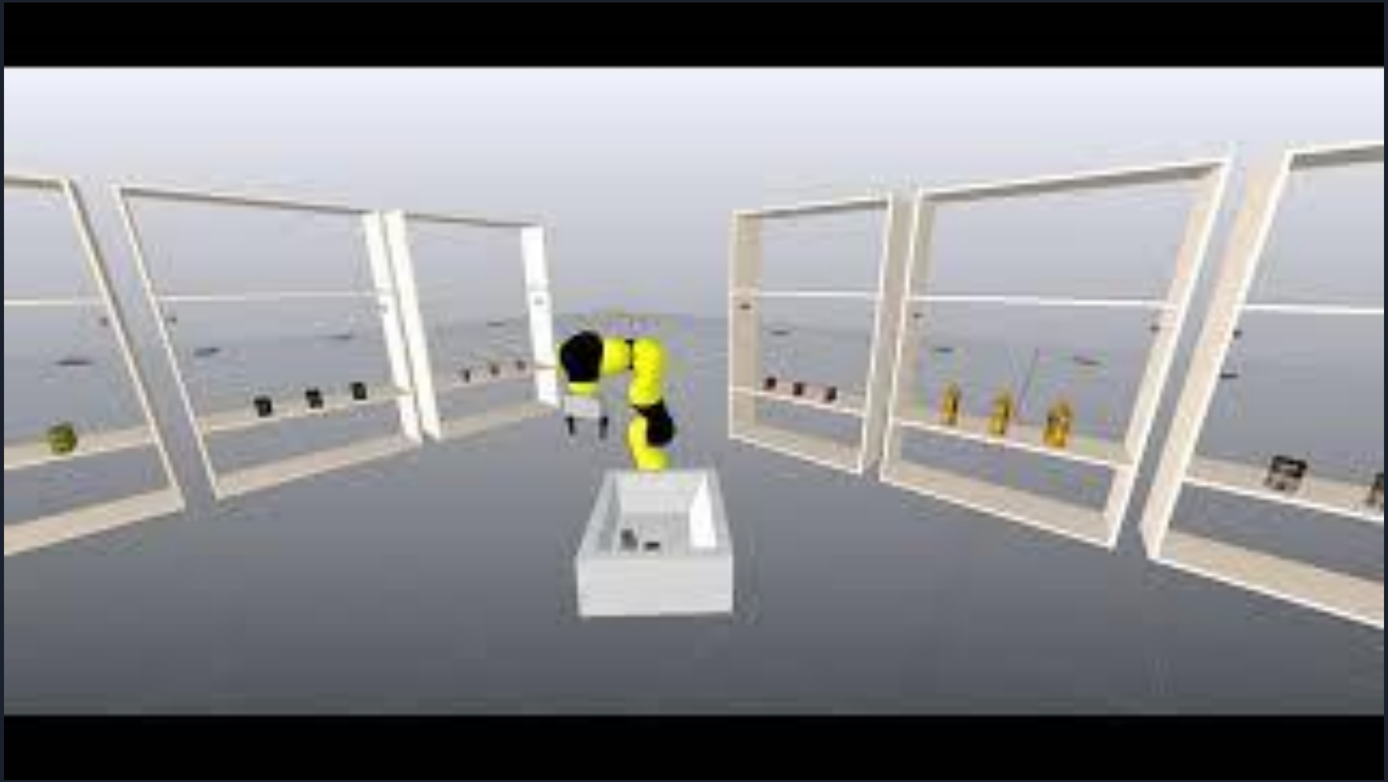
- Several environments created dynamically
- Rows and U-shaped
- Items generated on shelves
- One kind of item per shelf
- Shopping list parsed and given to the robot
- Item names mapped to shelves internally
- Simpler design (more items -> longer runtime)

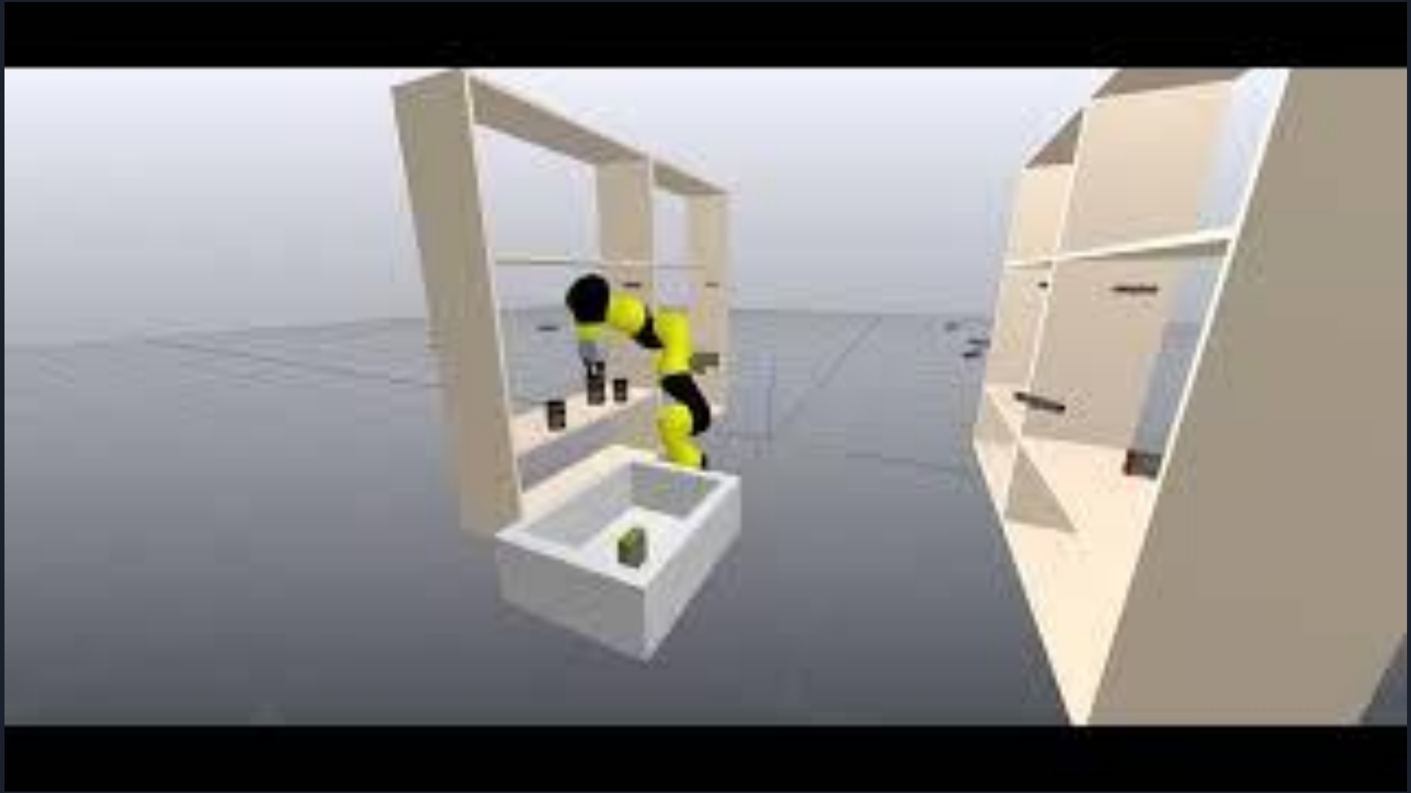




DEMO







Contributions

Emir: Motion Planning with DiffIK and Kinematic Trajectory Optimization, Mobile Base, Pick and Place, Task Planning

Efe: Environment Generation, Mobile Base, Pick and Place, GUI, TSP

Eren: Perception, Environment Generation, Pick and Place, TSP

Arda: Pick and Place, Perception, Mobile base, Environment Generation

Mert: Environment Generation, Pick and Place, Mobile Base

Project’s Benefits

- Automated collection of items from shelves
- Achieve shorter delivery times for customers
- Enable safer collection of items (robots can handle heavy items without risk)
- More accurate, reduce human errors (selecting wrong product or quantity)
- Extension in future - inventory management & collecting data about preferences
- Scalable - can handle growing demand with multiple robots

Technical Challenges and Successes

- Drake framework
- Kinematic trajectory optimization—inefficient, unreliable
- Differential IK—efficient, successful

Lessons Learnt

- Adapting to completely new tools (physics, math)
- Working around low documentation
- Out of our comfort zone - important for computer engineers
- Working as a team on an unfamiliar subject

Next steps

The project can be extended by:

- 4 Cameras per shelf is not feasible for monetary costs
- Placing cameras on the robot to reduce total camera count
- Add different grippers for better object coverage (suction gripper)
- Extend/improve supermarket environment
- Multiple robots
- Create grocery stock management system for employees
- Collecting data about customer preferences and implement business analysis

Thank you for listening