ShelfHelp: Empowering Humans to Perform Vision-Independent Manipulation Tasks with a Socially Assistive Robotic Cane

Shivendra Agrawal, Suresh Nayak, Ashutosh Naik, Bradley Hayes

University of Colorado Boulder

CAIRO Collaborative AI and Robotics Lab

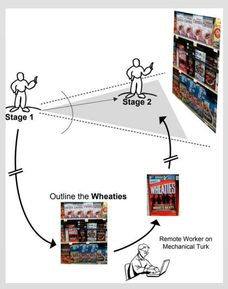
Motivation

- US market gets **~30,000** new products introduced each year, presenting a challenge for modern computer vision solutions.
- Working with sighted guide causes **lack of independence** and **loss of privacy** for blind and visually impaired (BVI).
- Similar densely packed items cause **poor tactile differentiability**.

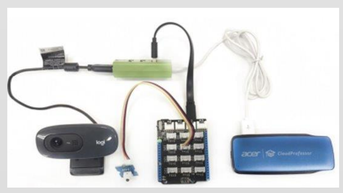
Finding a product is a hard visual problem

Related work

Manipulation Guidance
Bigham et al. [1]

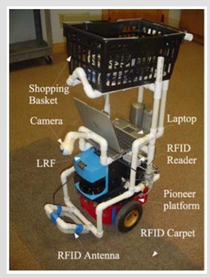


Product Detection
Feng et al. [2]



Limited # of products: 1329

Grocery Assistant
Gharpure et al. [3]



Existing works do not provide fine-grain manipulation guidance and require environmental augmentation

Our Solution

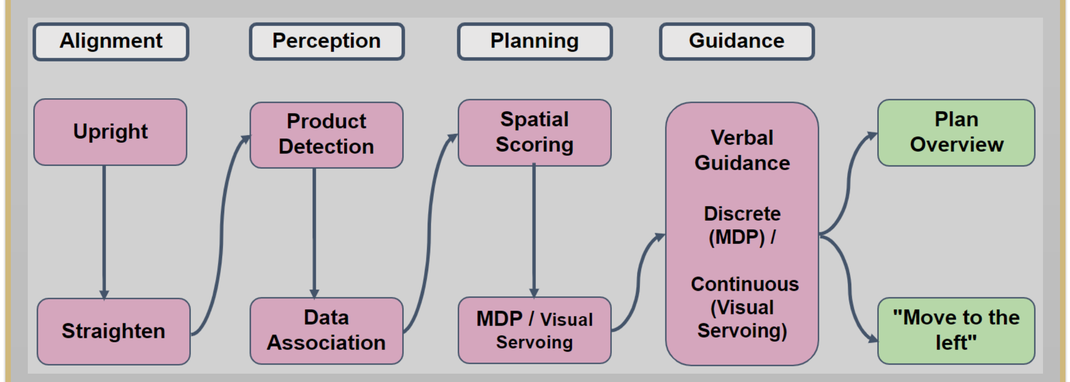
- **Domain:** Grocery shopping primarily consists of three main subtasks: navigation, product retrieval, and product examination. Our current work focuses on product retrieval.
- **Design consideration:** Repurpose navigational systems with sensing and compute.
- **Our contributions:** An end-to-end system that can assist with independent grocery shopping via

1) A modular and upgradeable 2-stage computer vision pipeline to locate desired products in a grocery setting.

2) A novel fine-grain manipulation guidance system that optimizes for guide time and the number of commands.

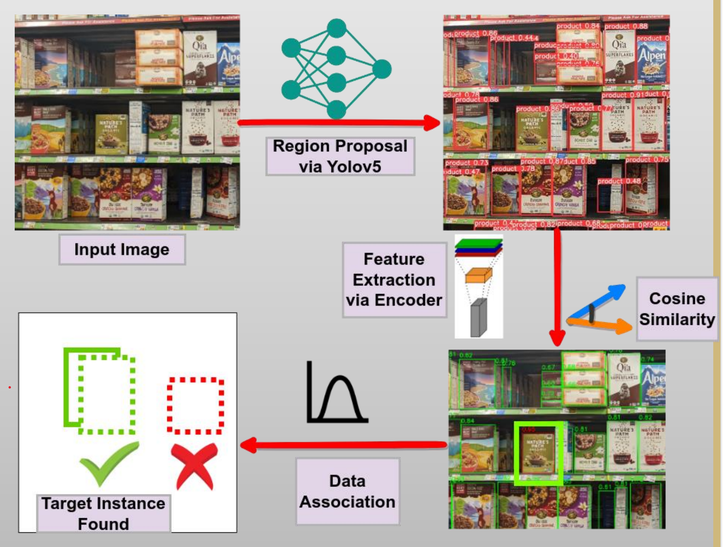
(Left) Our system used as a navigational assistance device.
(Right) Our system used as a manipulation assistance device.

Software Architecture



Product Detection

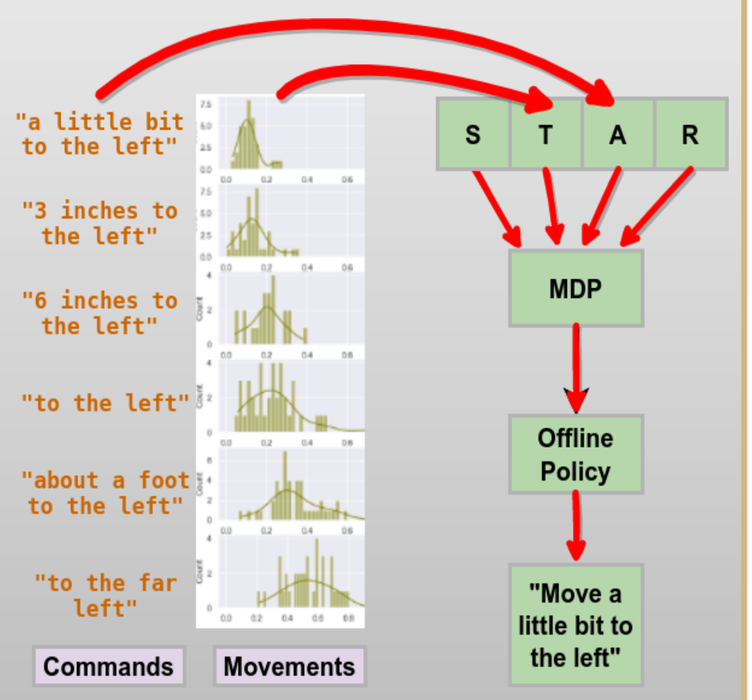
1. ShelfHelp employs a 2-stage method to detect the desired product.
2. YoloV5 network is used to give region proposals that are bounding boxes most likely to contain *any* product.
3. The features of these regions are compared against that of the desired target product. We use an Encoder for feature extraction.
4. The most similar region based on cosine similarity is selected.



Manipulation Guidance

1. We recorded a mapping of verbal instructions to human hand movement.
2. A Gaussian is fit that gives us a mean μ and standard deviation σ.
3. We formulate the guidance planning problem as a Markov Decision Process (MDP) where the

- Set of verbal instructions → Action space **(A)**
- μ and σ → Transition probability **(T)**
- Remaining distance to the goal → State **(S)**
- Reward **(R)** is designed to account for minimum number of command and legibility.

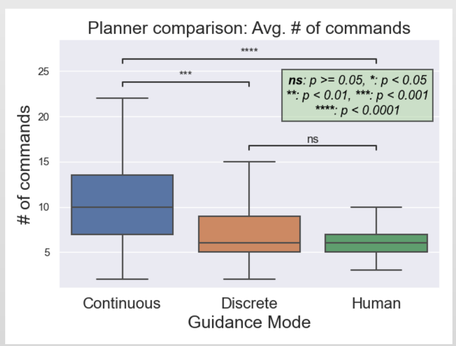


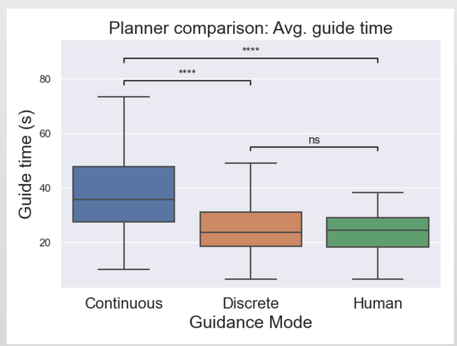
Study Design

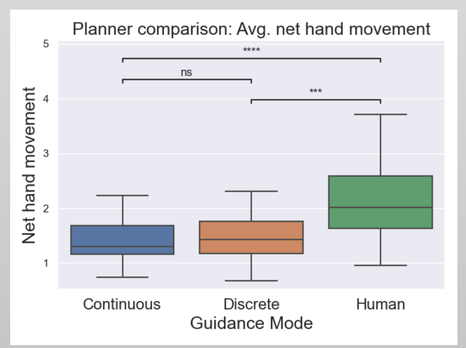
- We conducted an IRB-approved study with novice, blindfolded users (n=15) for system testing and validation.
- We compared the following planners -

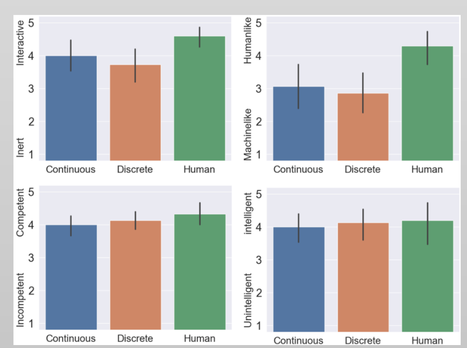
1. **Our planner (discrete planner)**
2. **Baseline planner (continuous planner)**
3. **Human guide (the gold standard)**

Results









1. **ShelfHelp outperformed the baseline planner in terms of # of commands and total guide time.**
2. **Matched human level performance on the above metrics.**
3. **Elicited positive response for confidence, intelligence, and competence**

References: [1] Jeffrey P Bigham, Chandrika Jayant, Andrew Miller, Brandyn White, and Tom Yeh. 2010. VizWiz:: LocateIt-enabling blind people to locate objects in their environment. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops. IEEE, 65–72.
[2] Chia-Hui Feng, Ju-Yen Hsieh, Yu-Hsiu Hung, Chung-Jen Chen, and Cheng-Hung Chen. 2020. Research on the Visually Impaired Individuals Shopping with Artificial Intelligence Image Recognition Assistance. In International Conference on Human-Computer Interaction. Springer, 518–531.
[3] Chaitanya P Gharpure and Vladimir A Kulyukin. 2008. Robot-assisted shopping for the blind: issues in spatial cognition and product selection. Intelligent Service Robotics 1, 3 (2008), 237–251.