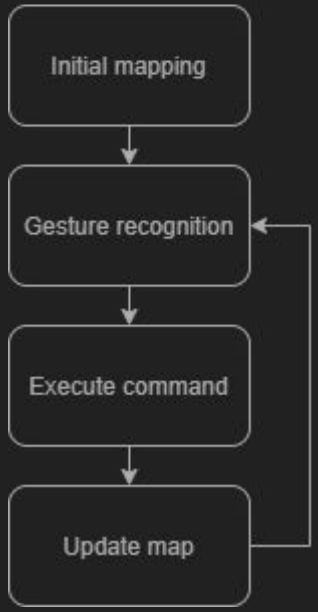
Team F2
Cookiebot - A Gesture Based Home Robot

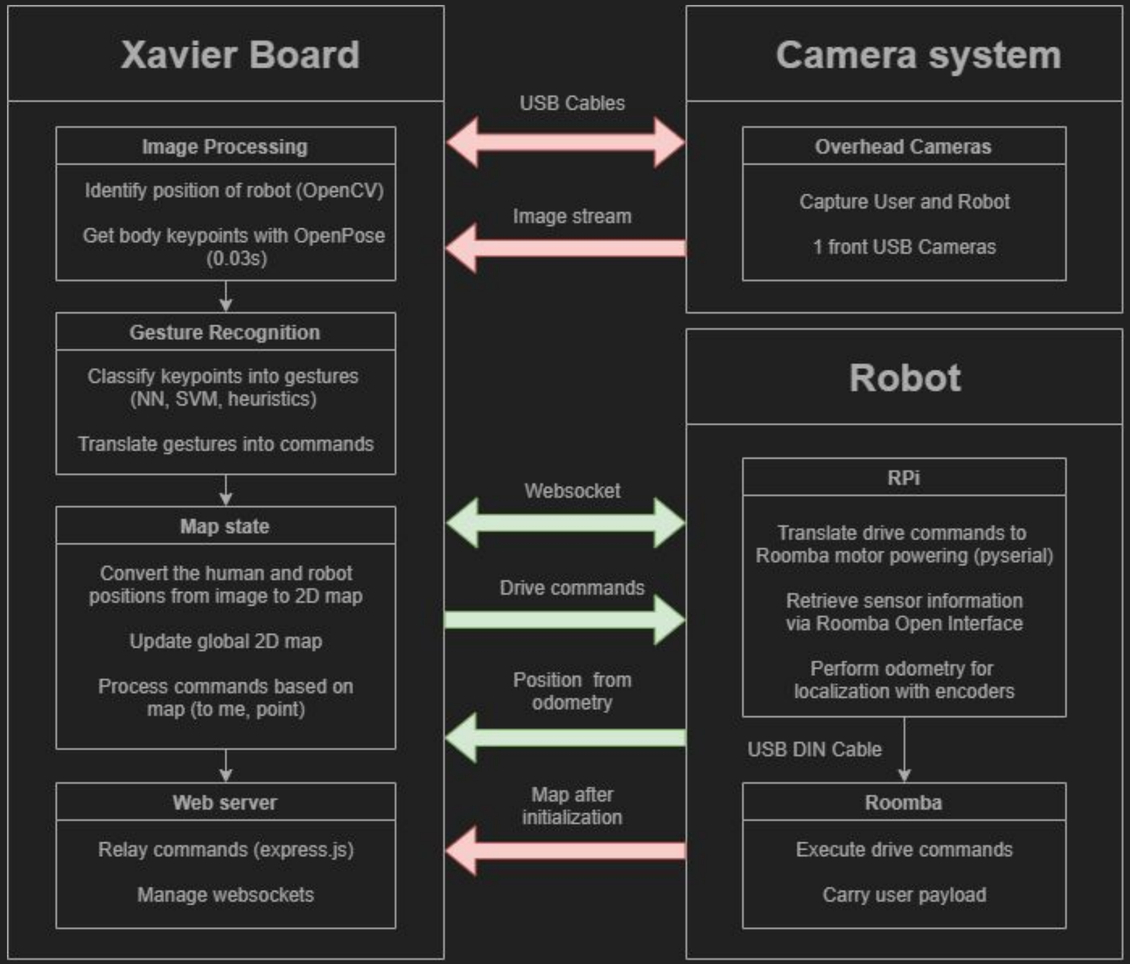
Seungmin Ha, Rama Mannava, Jerry Yu

Application Areas

- Home robot with intuitive gesture control
- Our target tasks:
  - Tele-operated robot control via gestures
  - Drive home to dock on command
  - Drive to a location the user points to
  - Drive to the user to deliver goods (cookies)
- Scope
  - One person - Limit complexity given time constraints
  - 8 gestures - 3 for tele-op, 1 gesture for stop, 1 gesture for go home, 1 gesture for going to user, 1 gesture for point ready, 1 gesture for going to pointed location
  - Will not avoid obstacles outside of those mapped during initialization

Block Diagram



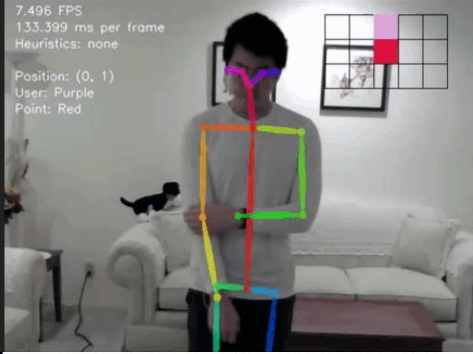


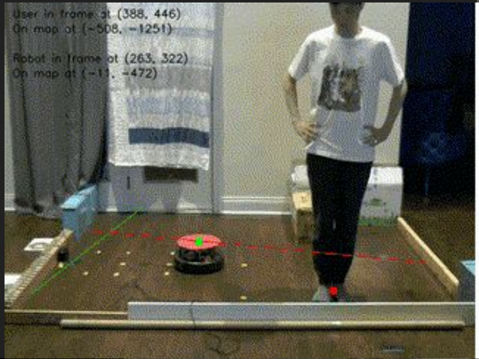
System Approach

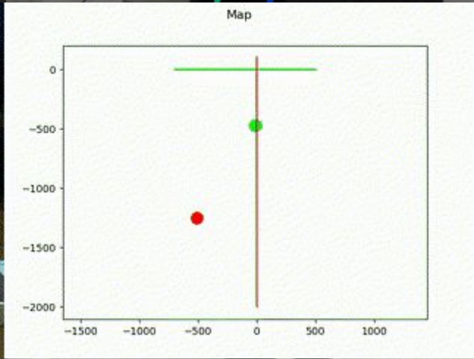
- CMU OpenPose (CNN) for pose keypoints
- Embedded GPU: NVIDIA Xavier Board
  - Run OpenPose at 27 FPS
  - Train and run models on GPU
  - Run expensive OpenCV operations on CPU
- Roomba
  - Communicate via pyserial and Roomba Open Interface
  - Access motors and sensors
- Custom data scripts
  - Collect and label data
  - Run system on streaming video for remote dev
  - Calculate metrics
- Visualizations

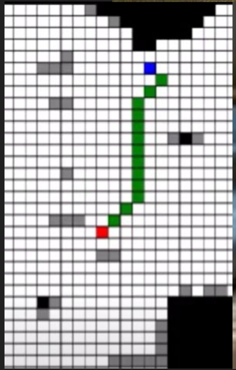
Complete Solution

- Gesture Recognition
  - Classify user gestures from OpenPose keypoints with heuristics, SVMs (teleop), NNs (point)
- 3D to 2D Mapping
  - Track where user and robot are with OpenPose and OpenCV
- Robot
  - Explore and map room
  - RasPi handles I/O and odometry
  - Drive to location with A*
- Demo
  - Visualizations
  - All 4 tasks in action









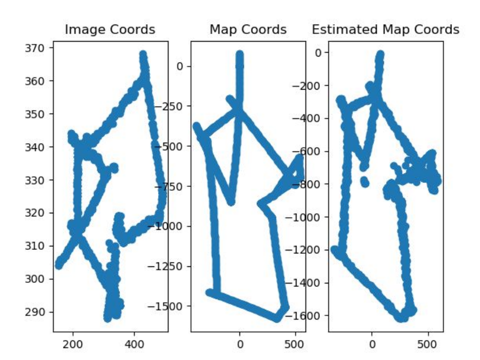
Metrics and Validation: Gesture Recognition

Tests assume user is in frame and face the camera
Tests are under different lighting conditions, are invariant based on position in room

Component	Metric	Method	Result
Teleop Classification (SVM)	> 90% accuracy (reasonable to use)	Tested on 76 gestures	Test accuracy: 90.8%
Point Classification (NN)	> 90% accuracy	Tested on 130 gestures	Test accuracy: 83.08%
Other Classification (Heuristics)	> 90% accuracy	Tested on 162 gestures	Test accuracy: 95.67%
All Gestures (Teleop, Point, Other)	> 90% accuracy < 10% misclassification (FP) < 5% unrecognized (FN)	Tested on 368 gestures	Accuracy: 90.2% Misclassification: 8.8% Unrecognized: 0.9%

Metrics and Validation: 3D to 2D Mapping

Component	Metric	Method	Result
User Position Accuracy	< 0.3 m drift on average (reasonable reaching distance)	Measure distance between the estimated position and the actual position over 3072 frames	0.04 m
Robot Position Accuracy	< 0.3 m drift on average (reasonable reaching distance)	Measure distance between the estimated position and the actual position over 3072 frames	0.12 m
3D to 2D Mapping Generation	< 60 sec per mapping (reasonable initialization phase overhead)	Time to generate mapping from 3072 frames of position data	4.8 sec



Metrics and Validation: Robot

Component	Metric	Method	Result
Robot Movement Accuracy	< 0.3 m drift on average (reasonable reaching distance)	Measure distance between the goal position and the actual position after 5 runs of movement	0.048 m
Robot Movement Speed	0.05 m/s to 0.2 m/s on average, safety and efficiency	Measure the time between the command and the movement completion of 10 gestures (~5m of travel each)	0.11 m/s
Robot Battery Life	> 2 hrs operation without recharging	Make the robot perform drive function every ten minutes, without recharging	Passed
Room Coverage	> 90% map accuracy	Count inaccurately represented 5 cm square cells from a 2D map	95.6%

Metrics and Validation: Performance

Component	Metric	Method	Result
Gesture Recognition Time (Xavier)	< 500 ms	Time from frame capture to recognition for 129 gestures	Average time: 56 ms Longest time: 261 ms
3D-2D Mapping Processing Time (Xavier)	< 100 ms	Time to calculate map positions from frame for 3072 frames	Average time: 16 ms Longest time: 166 ms
Robot Path Planning (Xavier)	< 1000 ms	Measure time for the robot to calculate a path to the goal after 6 runs	Average time: 230 ms Longest time: 780 ms
Robot RPi - Server Transmission (Xavier and RPi)	< 100 ms	Measure for 100 sends and receives	Average time: 3.5 ms Longest time: 13.5 ms
Robot Execution Time (RPi)	< 15ms	Measure 100 cycles from receiving command to execution	Average time: 13.7 ms Longest time: 16.2 ms
Total Time	< 1.9s (Google Home)	max(gesture, mapping) + path+ transmission + execution	Average time: 0.30 sec Longest time: 1.07 sec

Design Tradeoffs

- AWS GPU vs Xavier
  - Big GPU with network bandwidth vs GPU on local network
- Data Collection Time vs Accuracy
  - Point
    - Trained and tested on 1 hr of data
    - Trade accuracy for time to collect and label data
- Speed vs Accuracy
  - 3D to 2D Mapping
    - Frame dropping increases speed and decreases accuracy; ultimately fast enough
    - Trade accuracy for speed
  - Robot
    - Limiting robot speed improves odometry accuracy and user safety; lower efficiency
    - Trade speed for accuracy

Project Management