A Viewpoint on Perception, Planning, and Control

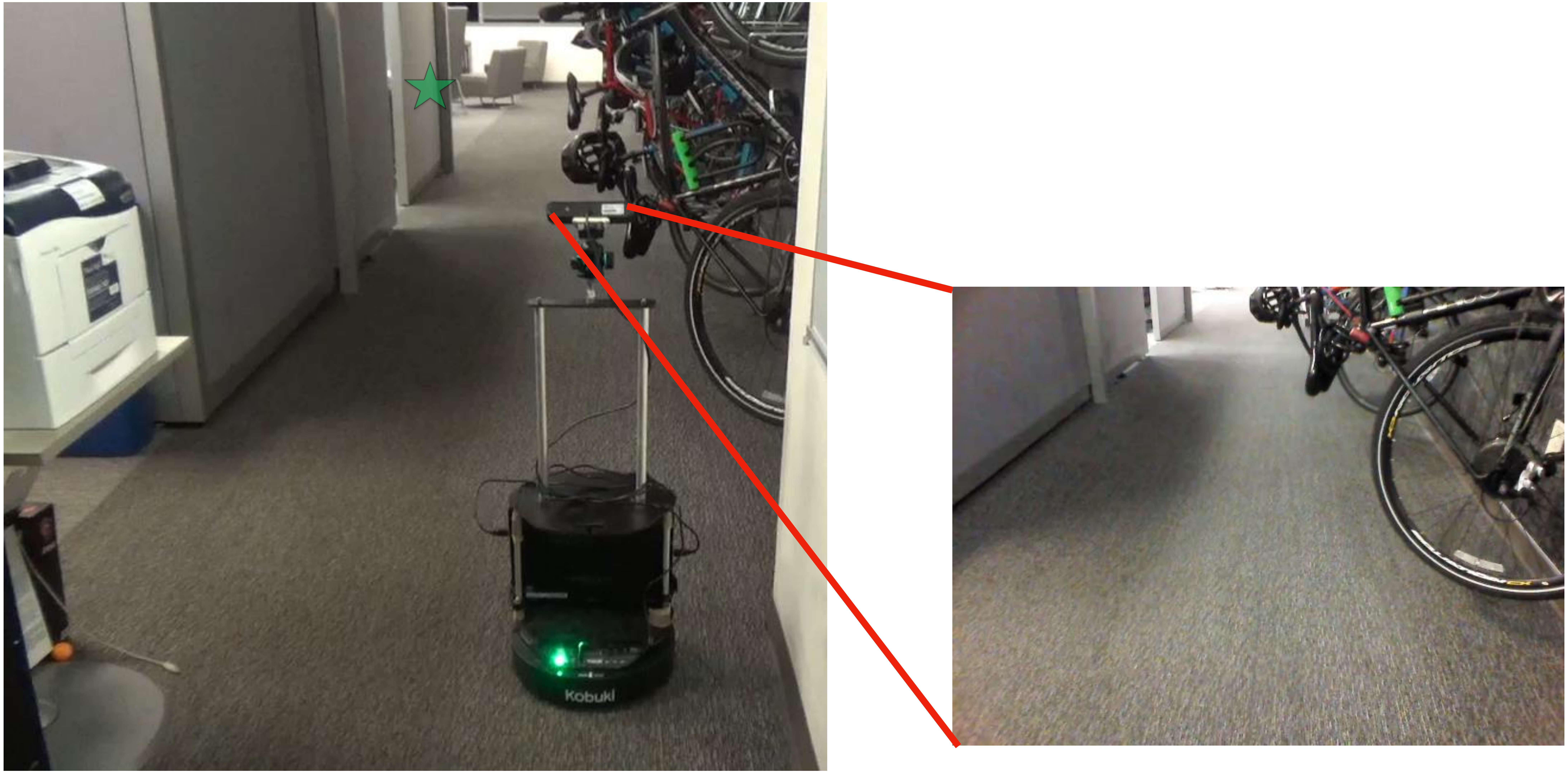
Claire Tomlin
Somil Bansal, Varun Tolani, Aleksandra Faust, Jitendra Malik

Hybrid Systems Laboratory

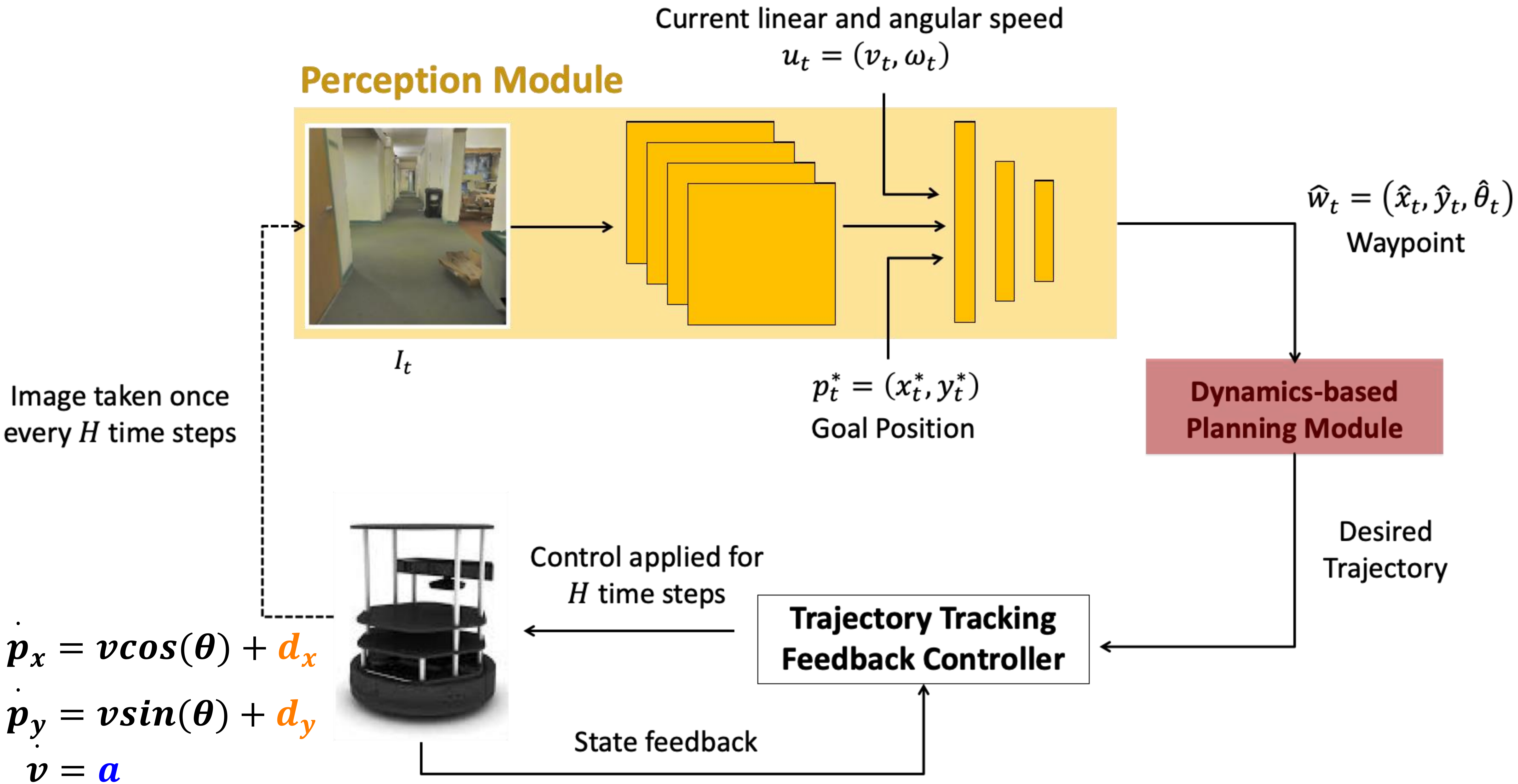
THE UNIVERSITY OF CALIFORNIA BERKELEY 1868

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

How to efficiently navigate an autonomous system with a monocular RGB camera to a goal in an *a priori* unknown environment?

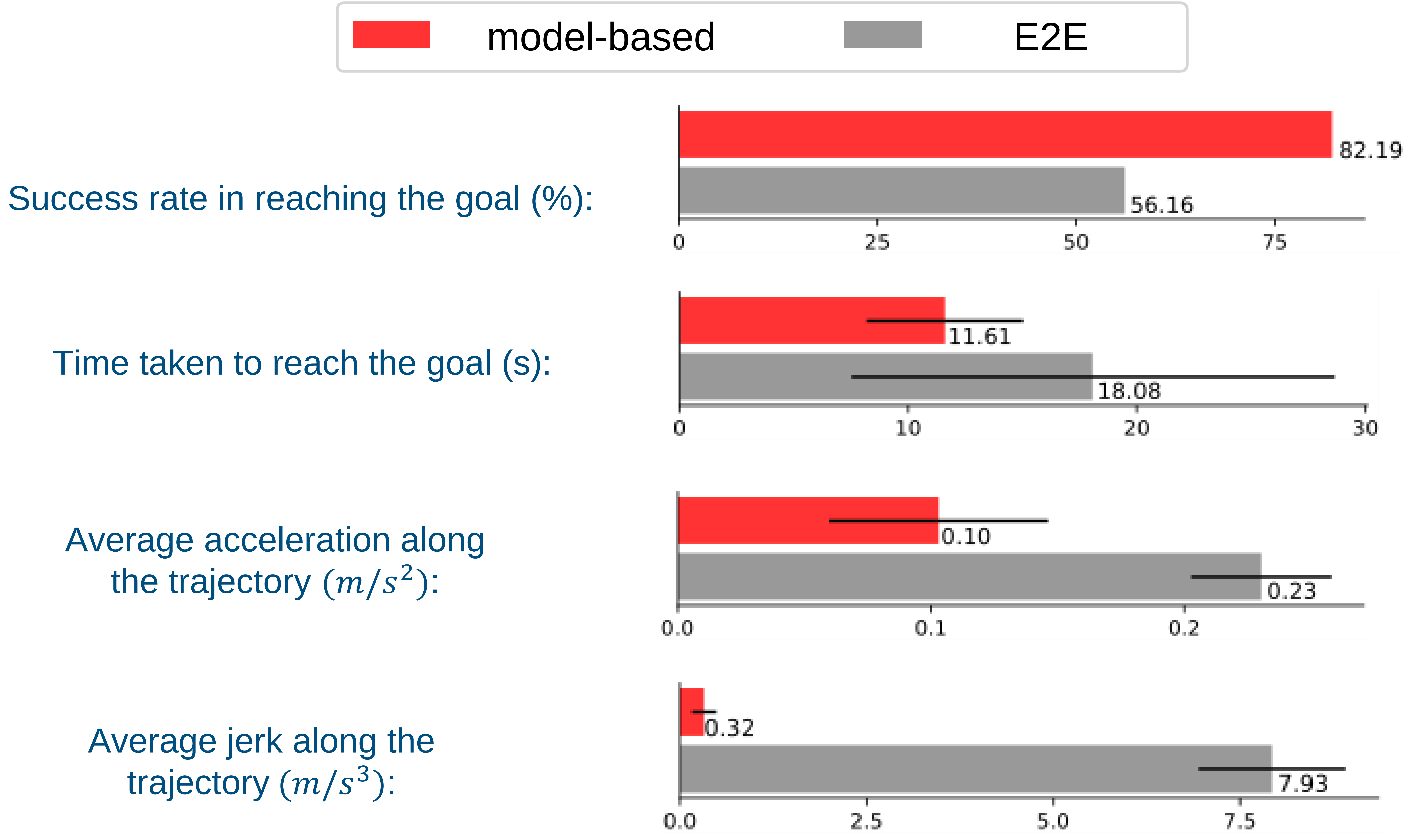


[Bansal, Tolani, Gupta, Malik, Tomlin, CoRL 2019]



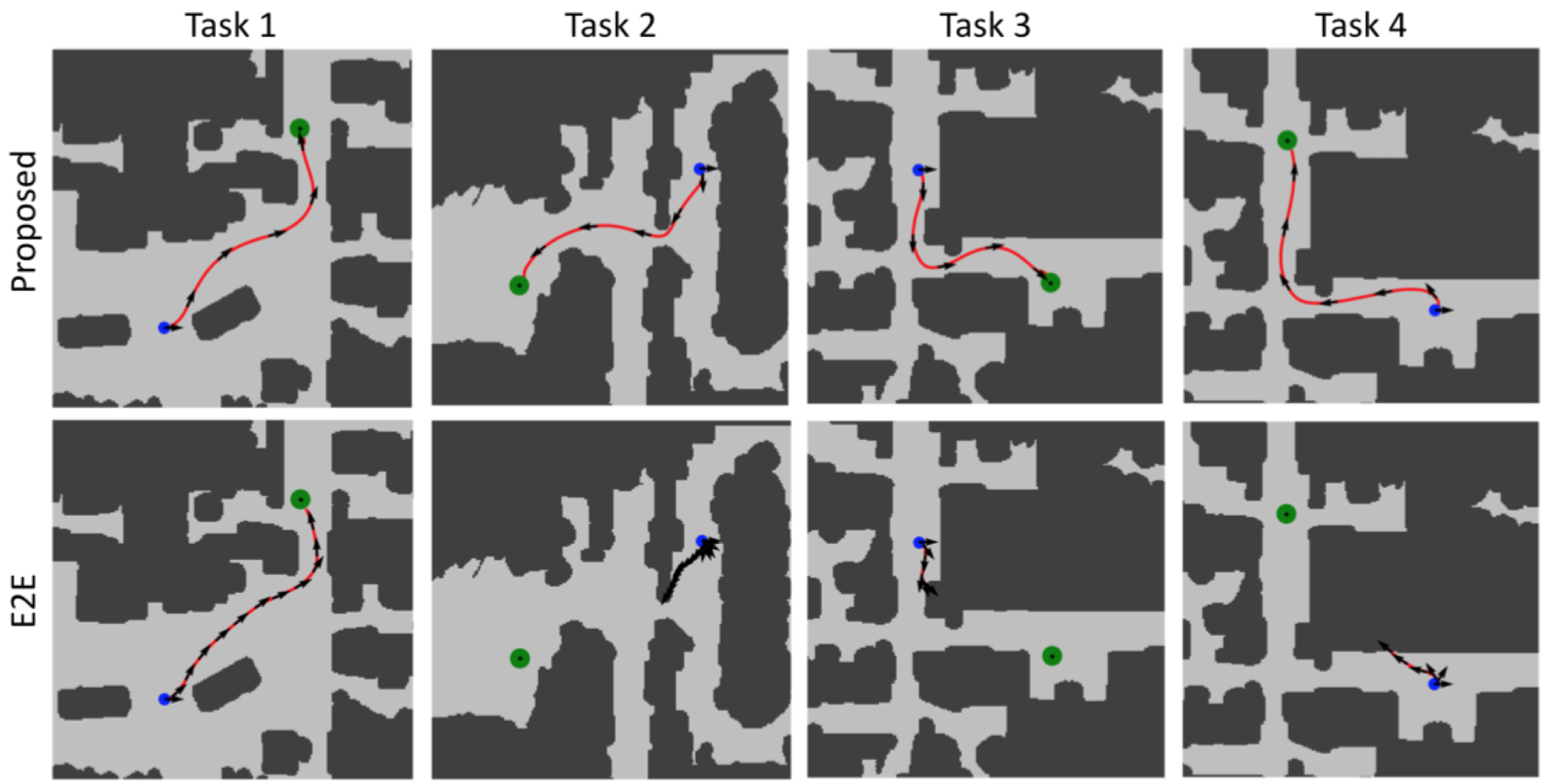
$$\dot{p}_x = v\cos(\theta) + d_x$$
$$\dot{p}_y = v\sin(\theta) + d_y$$
$$\dot{v} = a$$
$$\dot{\theta} = \omega$$

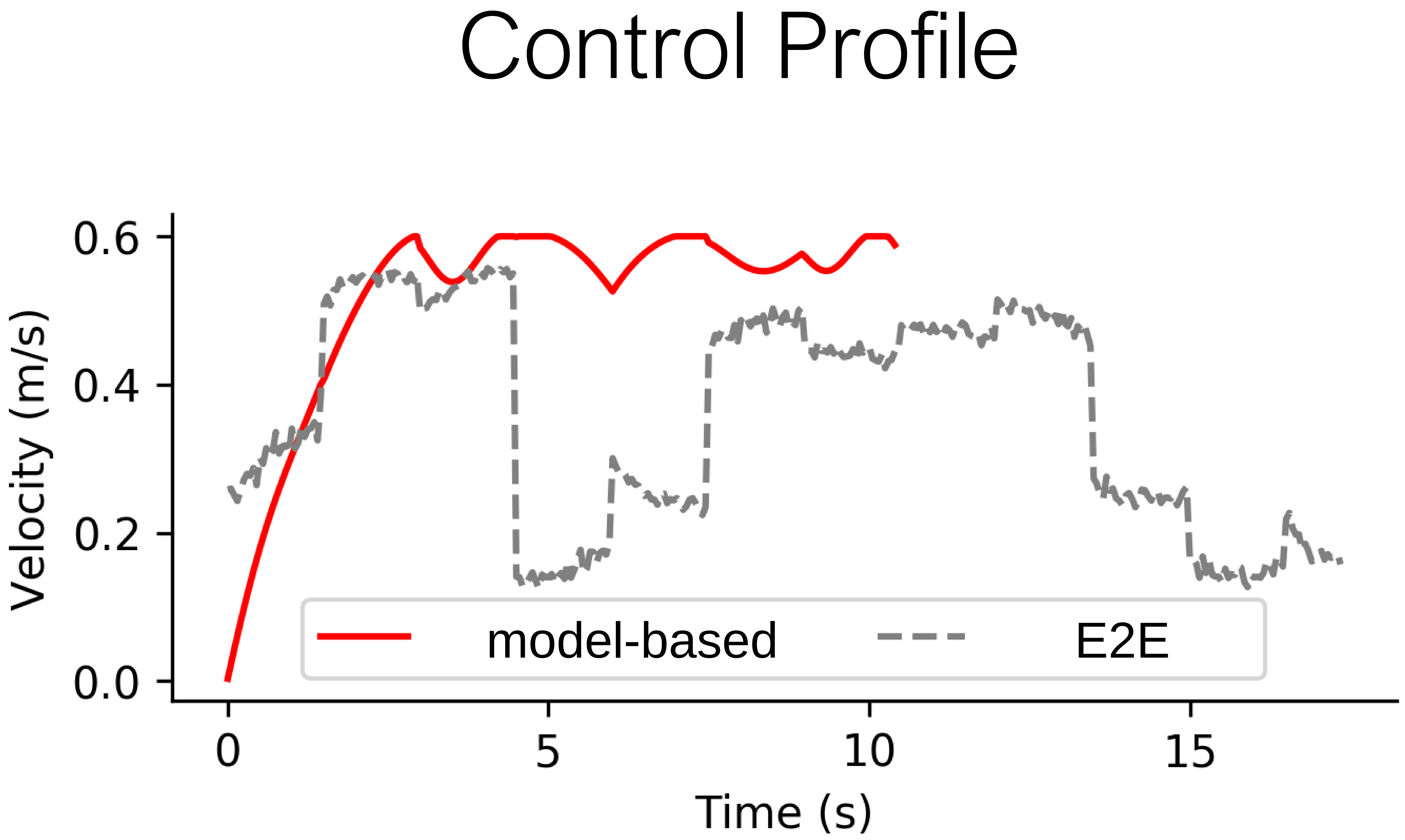
[Bansal, Tolani, Gupta, Malik, Tomlin, CoRL 2019]


model-based
E2E
Success rate in reaching the goal (%):
82.19
56.16
0
25
50
75
Time taken to reach the goal (s):
11.61
18.08
0
10
20
30
Average acceleration along the trajectory (m/s^2):
0.10
0.23
0.0
0.1
0.2
Average jerk along the trajectory (m/s^3):
0.32
7.93
0.0
2.5
5.0
7.5

Agent	Input	Success (%)	Time taken (s)	Acceleration (m/s^2)	Jerk (m/s^3)
Expert	Full map	100	10.78 ±2.64	0.11 ±0.03	0.36 ±0.14
LB-WayPtNav (our)	RGB	80.65	11.52 ±3.00	0.10 ±0.04	0.39 ±0.16
End To End	RGB	58.06	19.16 ±10.45	0.23 ±0.02	8.07 ±0.94
Mapping (memoryless)	Depth	86.56	10.96 ±2.74	0.11 ±0.03	0.36 ±0.14
Mapping	Depth + Spatial Memory	97.85	10.95 ±2.75	0.11 ±0.03	0.36 ±0.14

Success Rate

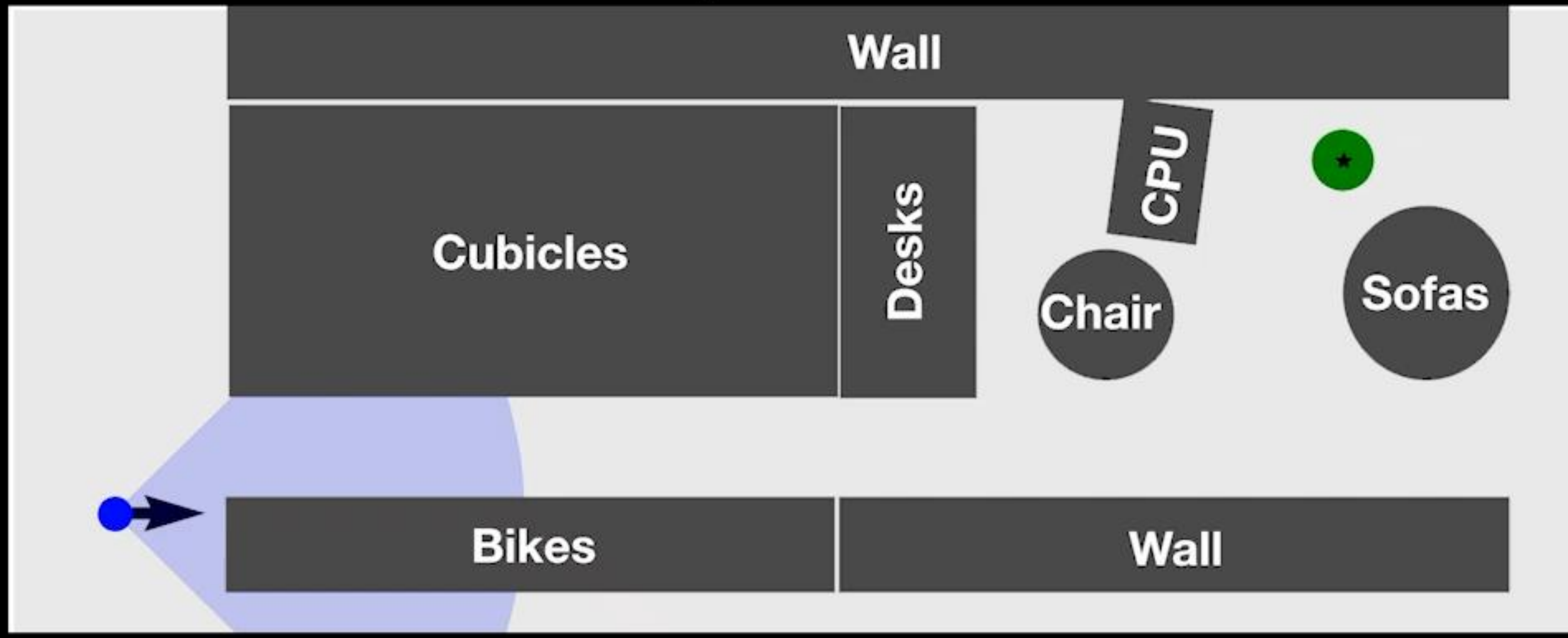




Control Profile
Velocity (m/s)
0.0
0.2
0.4
0.6
0
5
10
15
Time (s)
model-based
E2E

First Person View

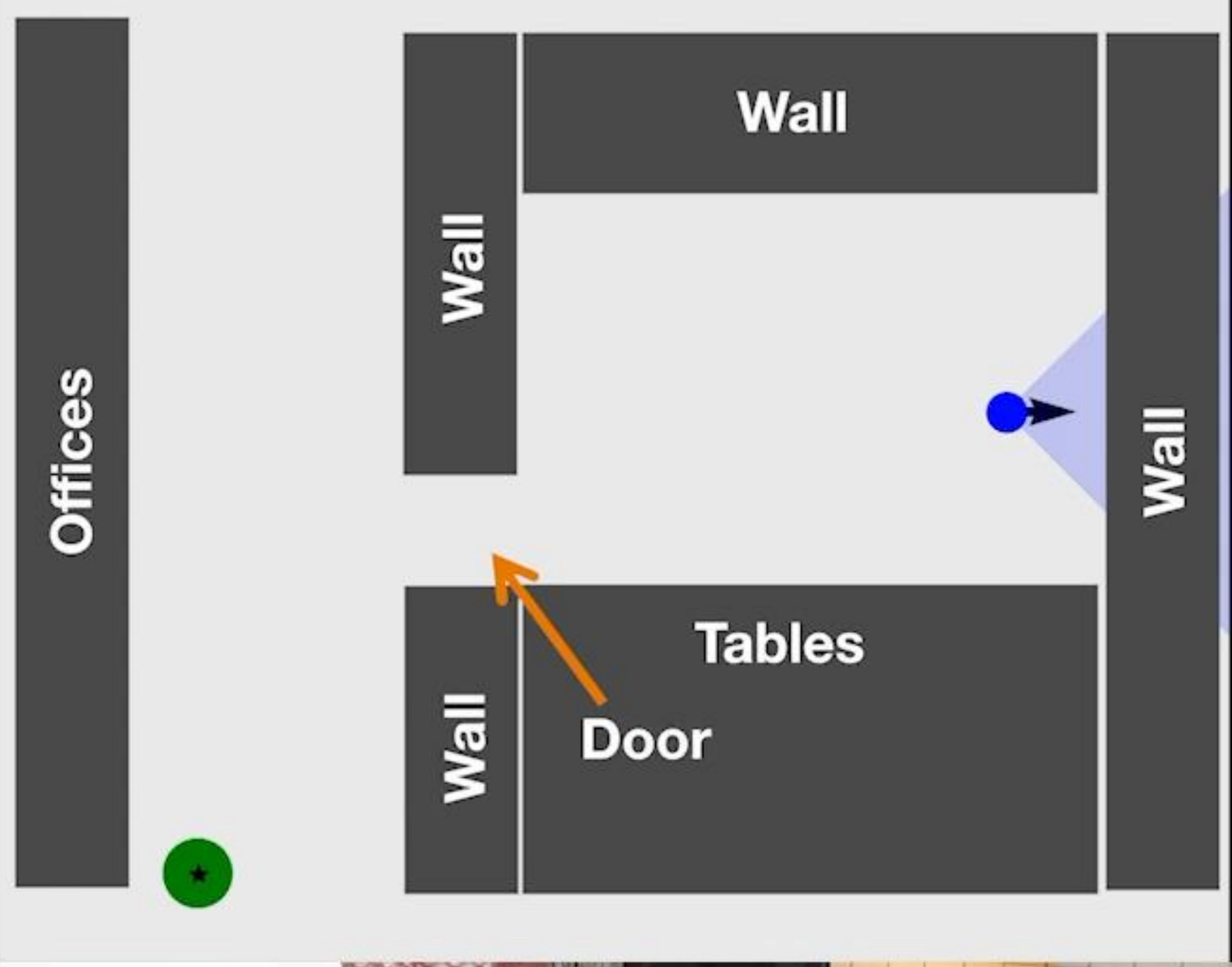
Top View



Third Person View

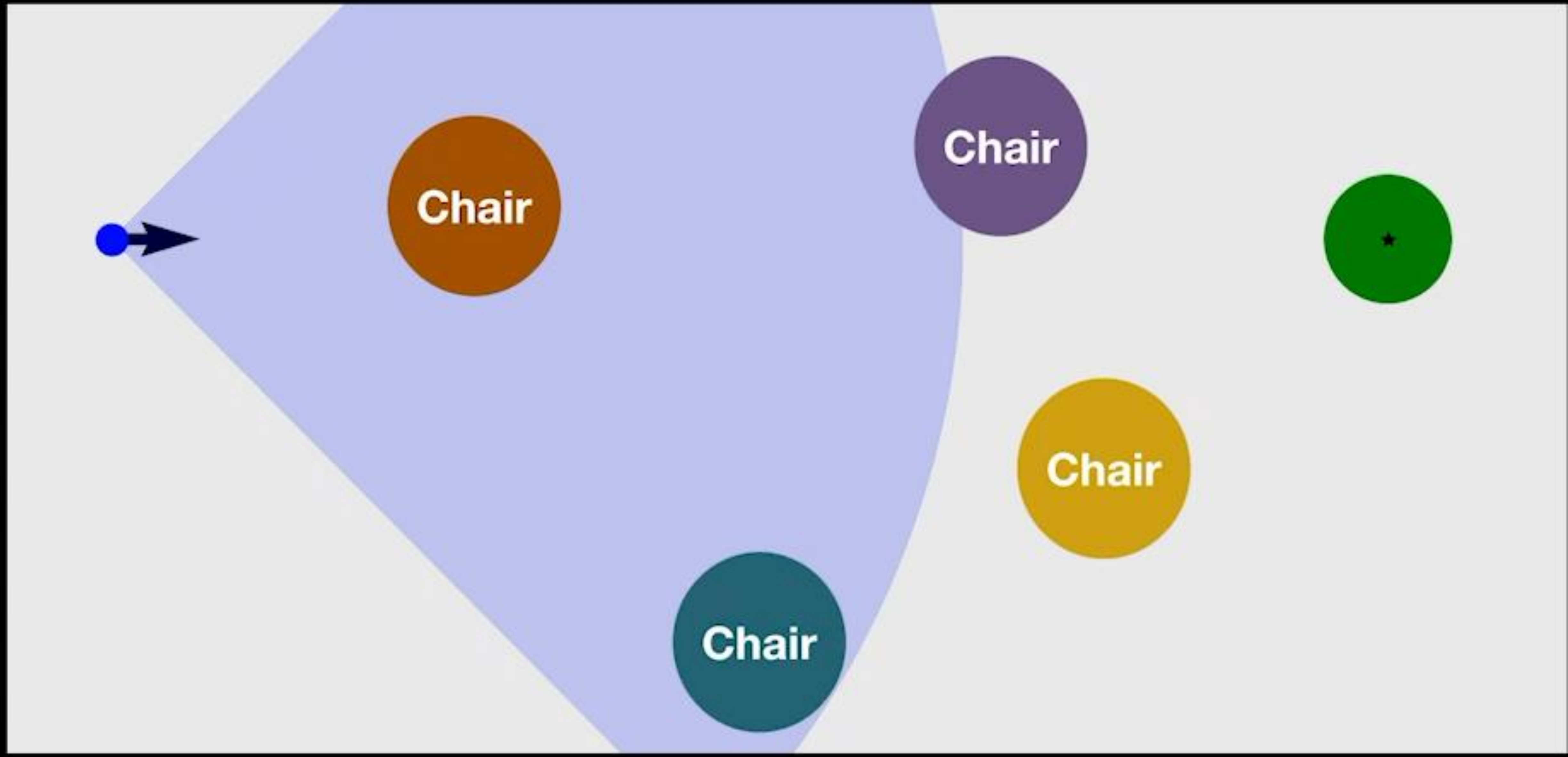
First Person View

Top View



Third Person View

Top View



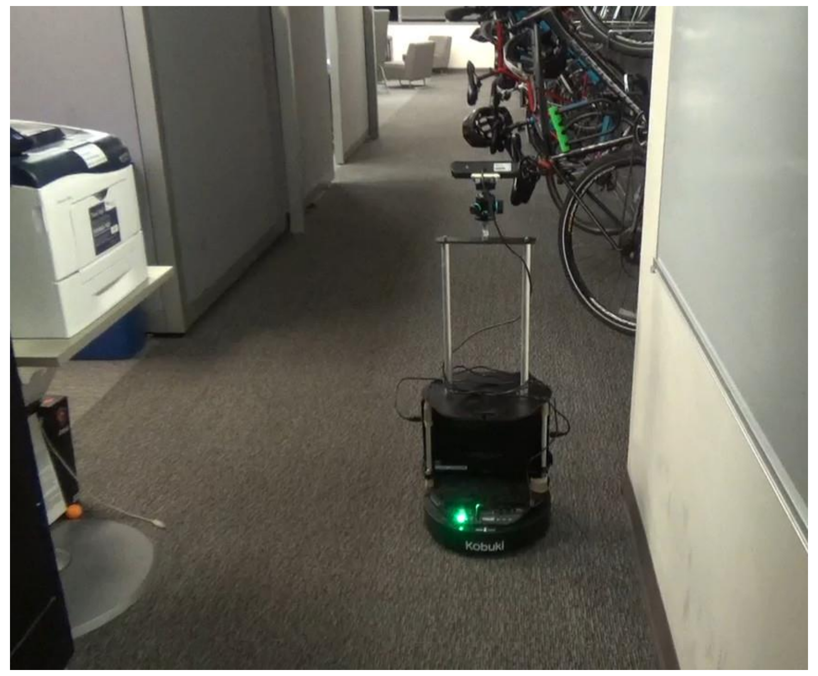
First Person View

Third Person View

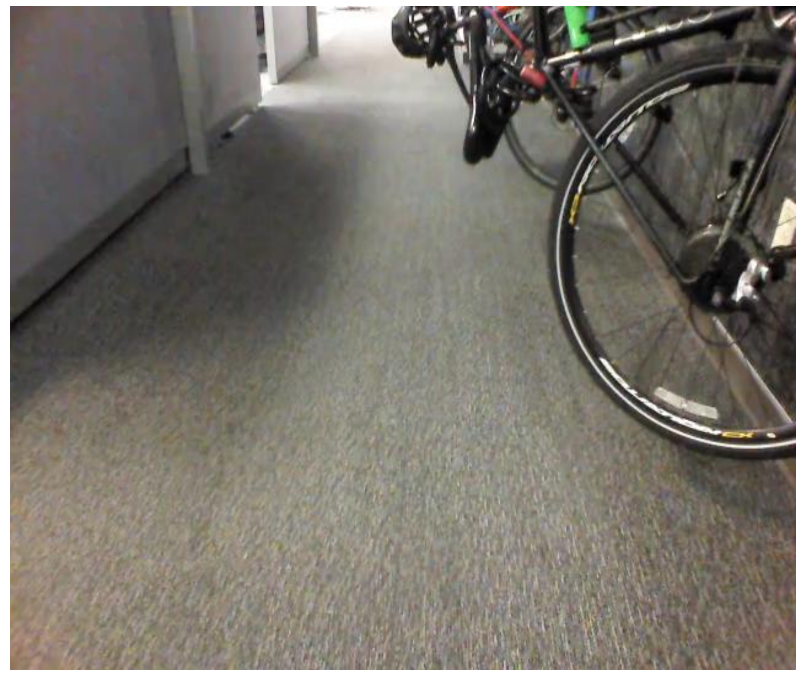
Agent	**Input**	**Success** (%)	**Time taken** (s)	**Acceleration** (m/s^2)	**Jerk** (m/s^3)
LB-WayPtNav (our)	RGB	95	22.93 ±2.38	0.09 ±0.01	3.01 ±0.38
End To End	RGB	50	33.88 ±3.01	0.19 ±0.01	6.12 ±0.18
Mapping (memoryless)	RGB-D	0	N/A	N/A	N/A
Mapping	RGB-D + Spatial Memory	40	22.13 ±0.54	0.11 ±0.01	3.44 ±0.21

Some lessons learned

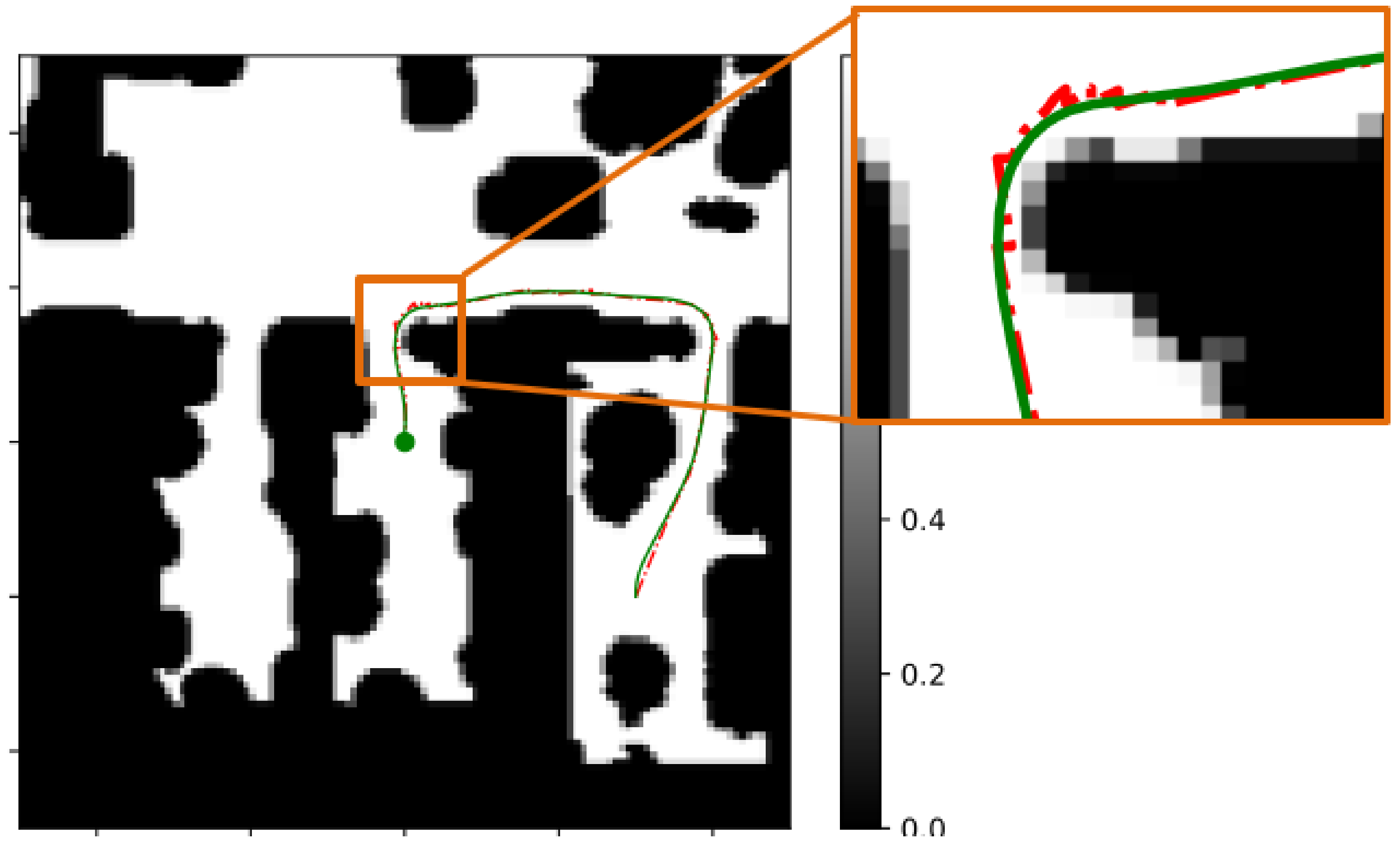
- Data representation is important



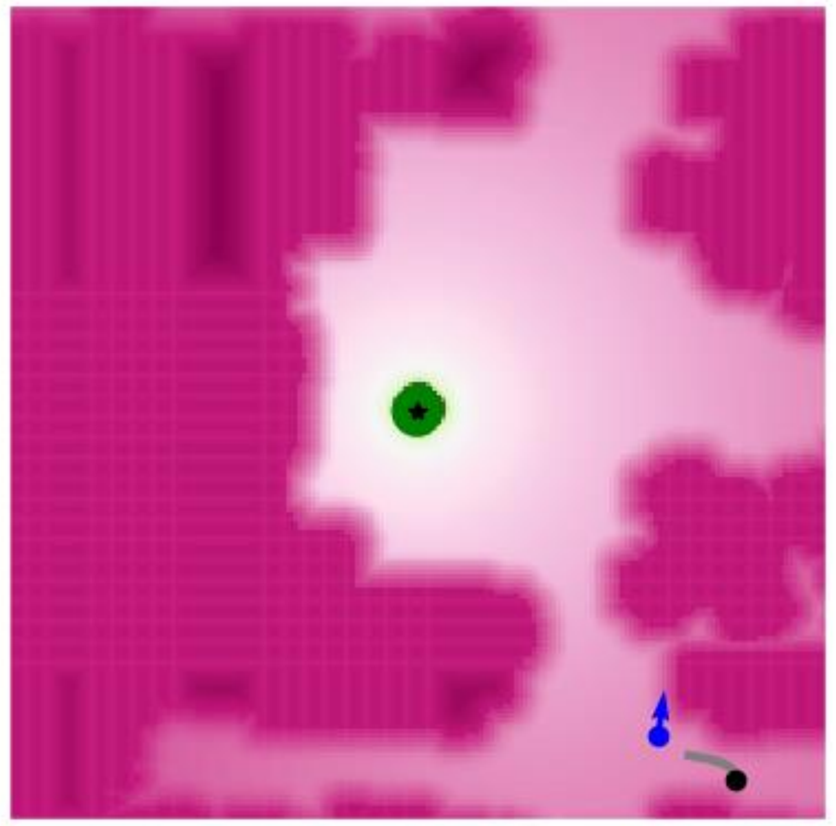
vs

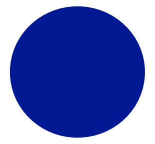


- Optimal control can be too optimal

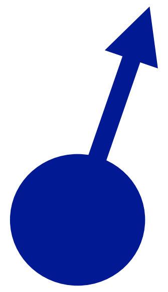


- Waypoint representation





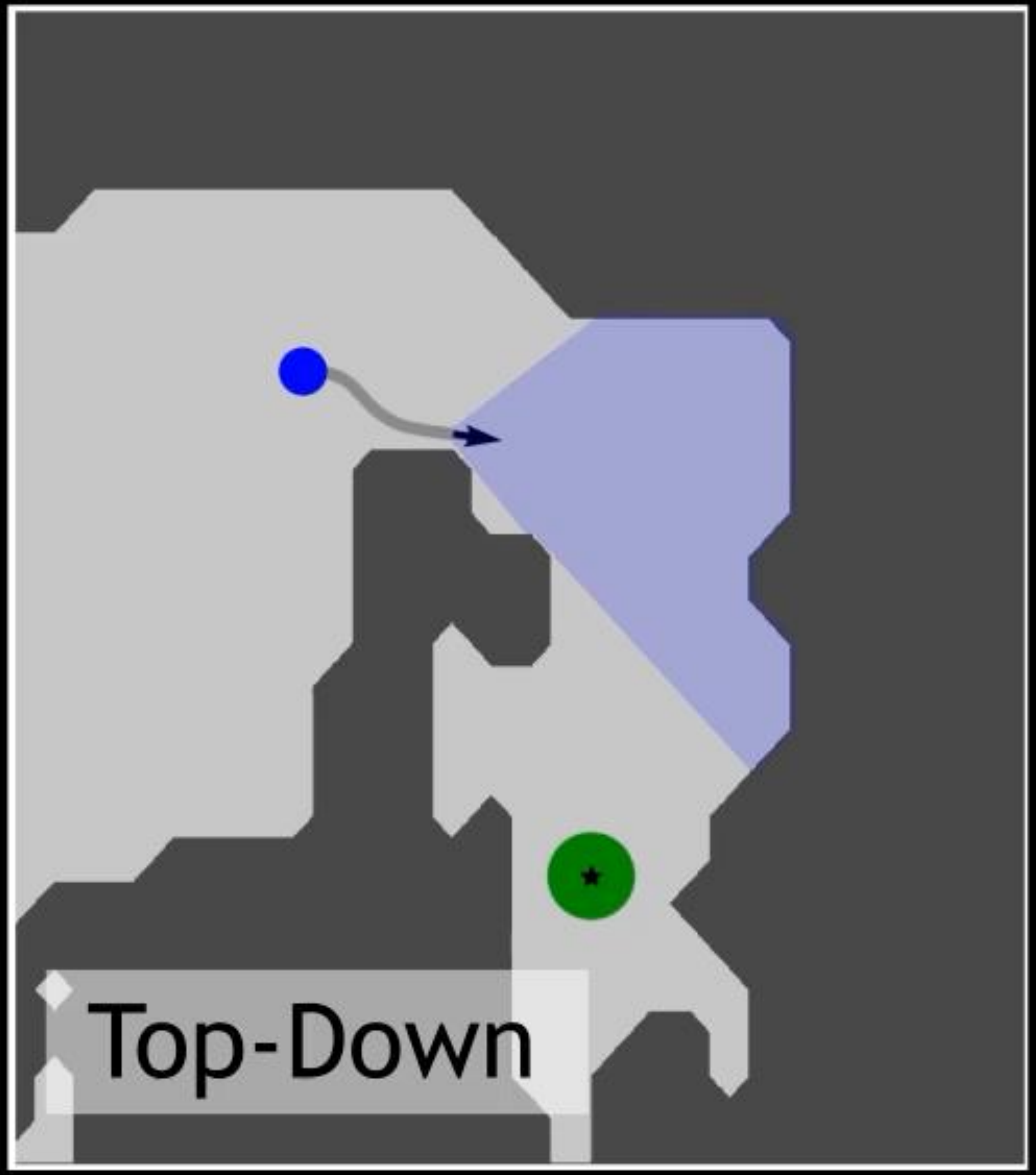
vs



More lessons learned

- Building on existing NN architectures
- Image and perspective distortions during training
- RL on supervised learning


Goal
Kobuki
Third-Person POV


Top-Down

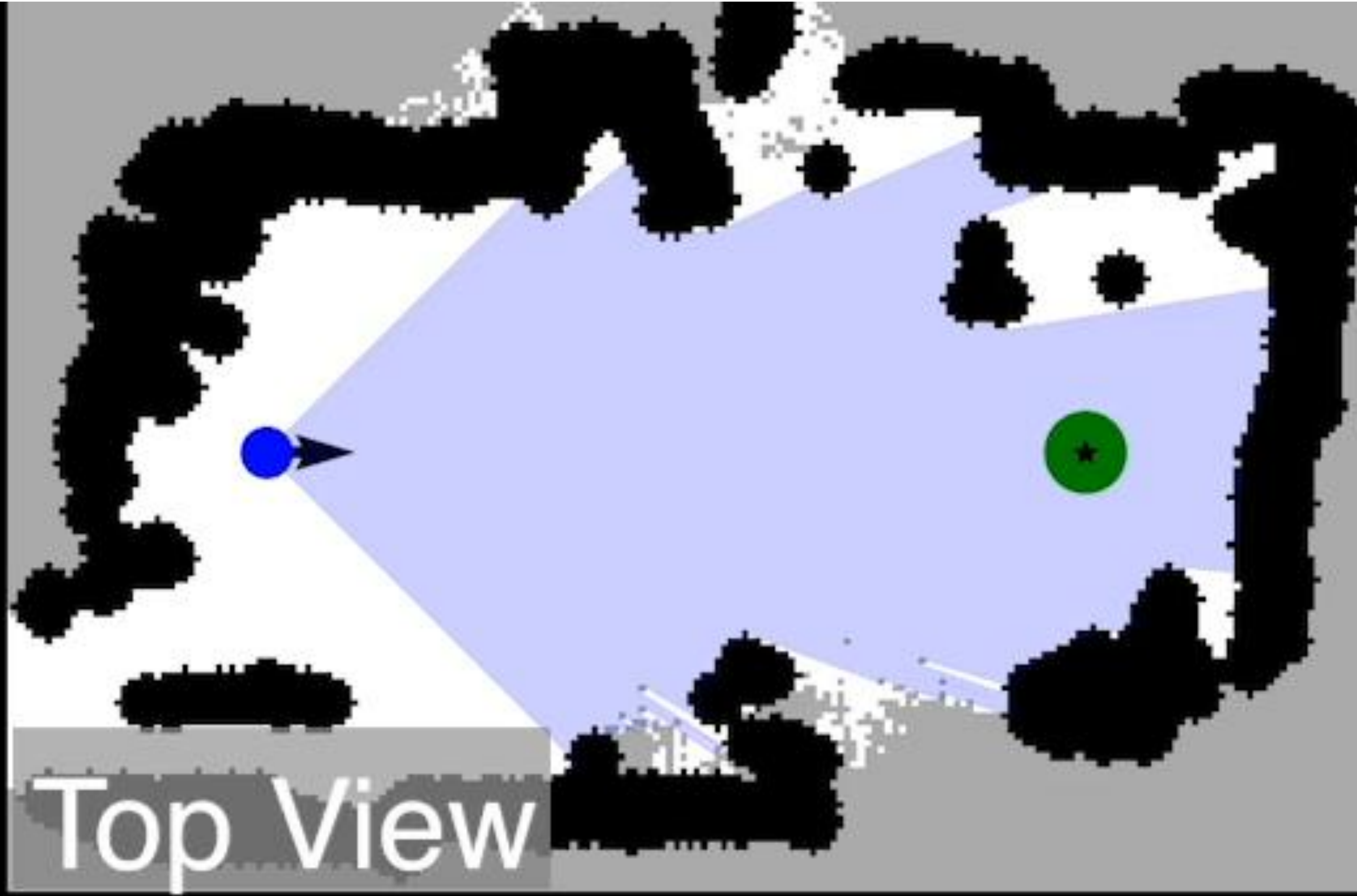

Robot POV

Safety Challenges

- Monitor: Is the image data in the training distribution?
- What is the uncertainty around the output of the perception module?
- How should this uncertainty affect the planning and control?
- More complex environments?

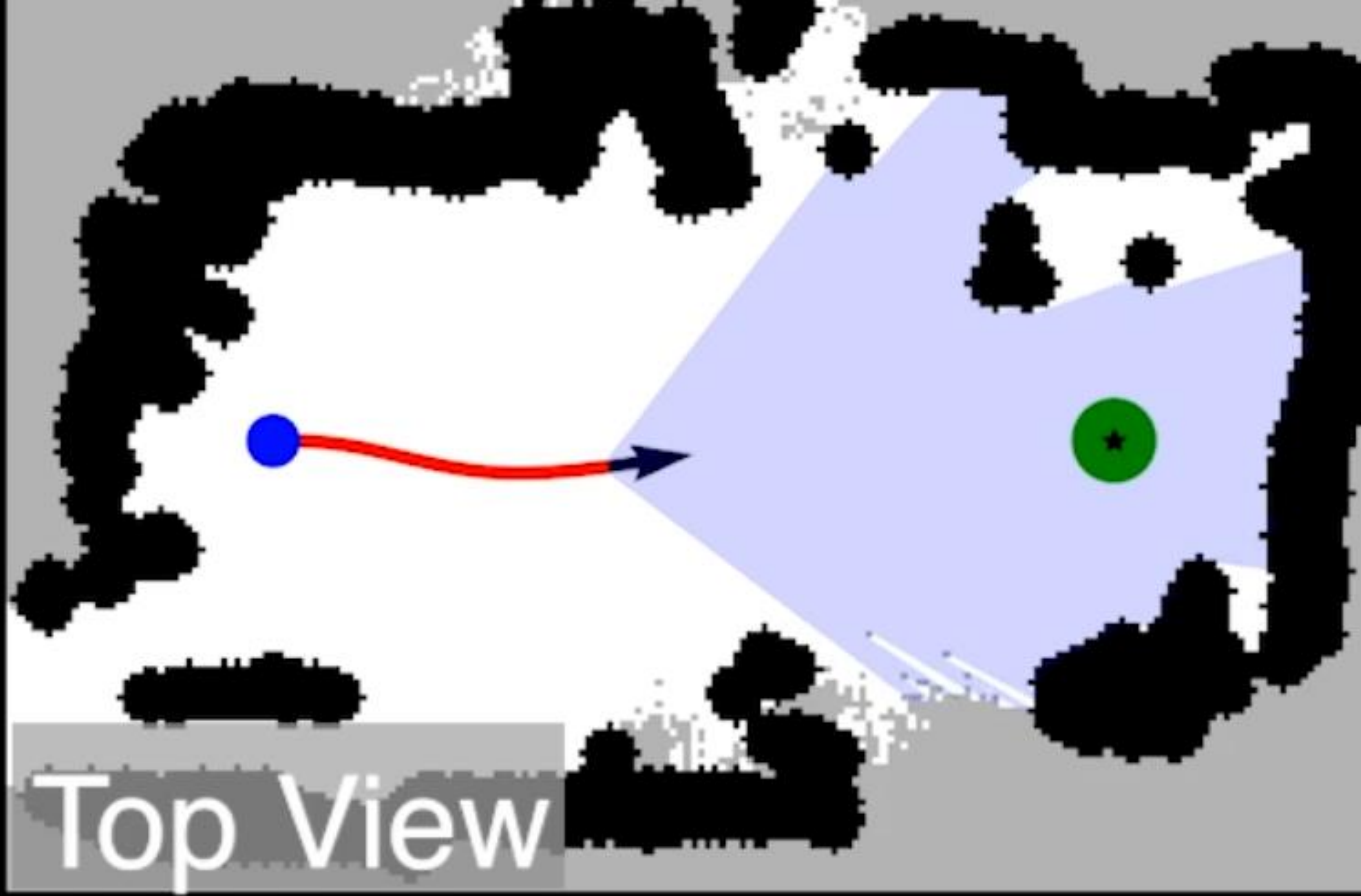
LB-WayPtNav

Experiment 1 (1x)



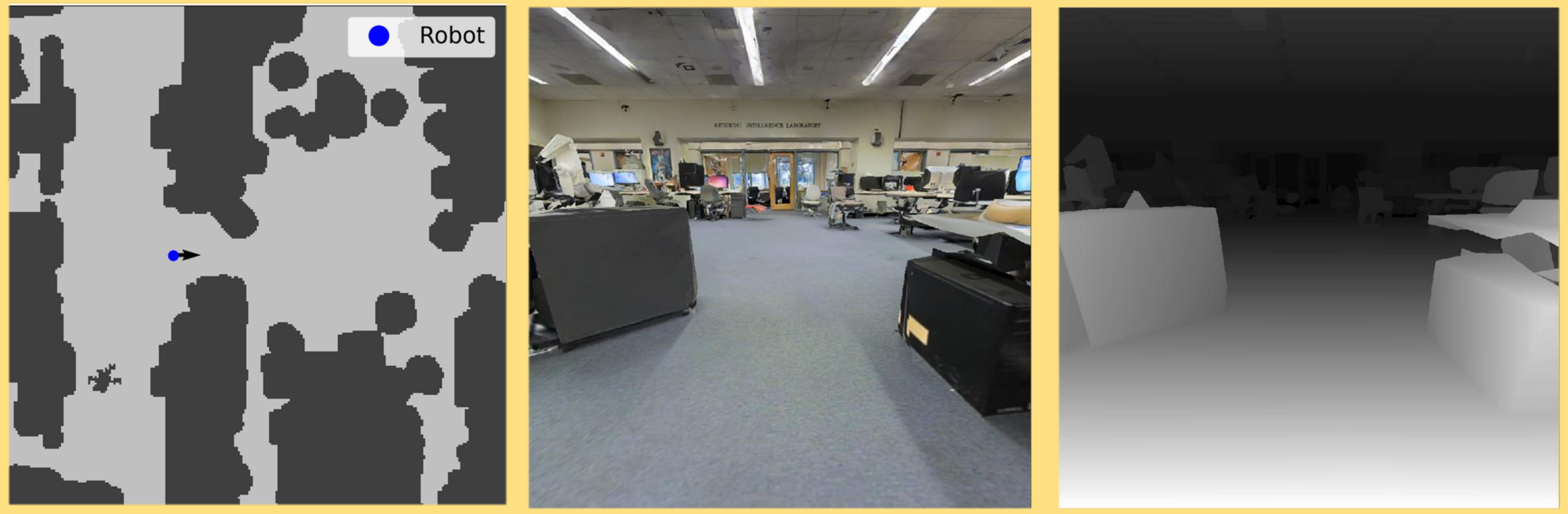
LB-WayPtNav

Experiment 1 (1x)

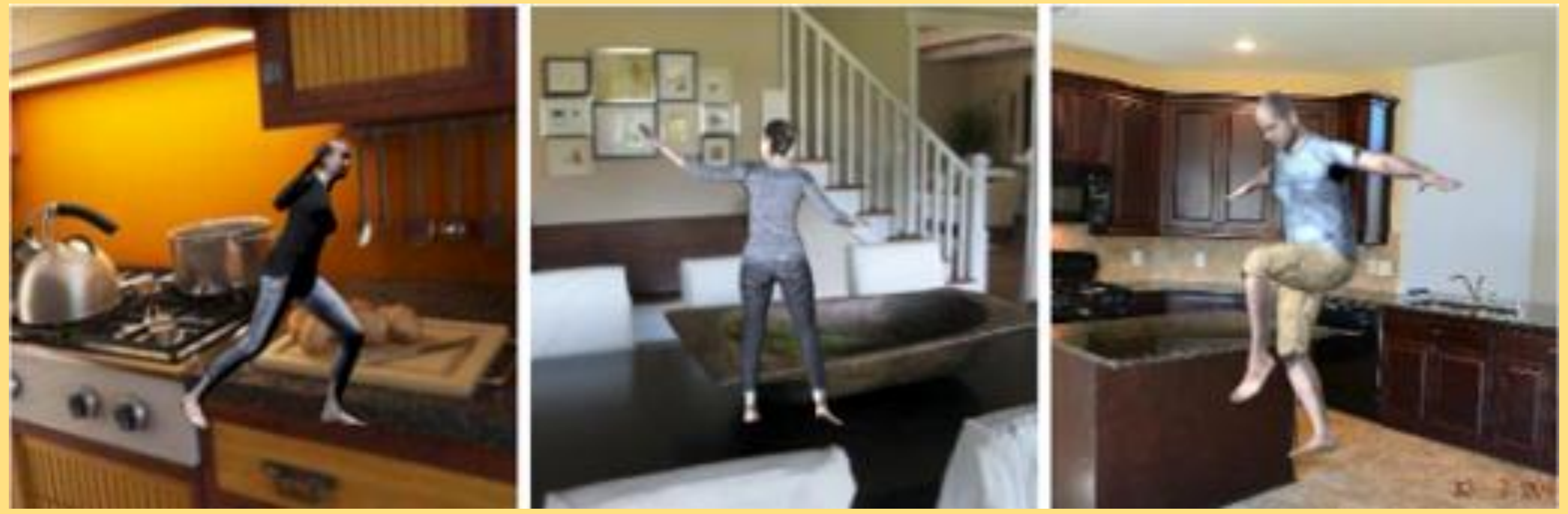


HumANav

SD3DIS



SURREAL



HumANav

Robot State

$[x, y, \theta]$

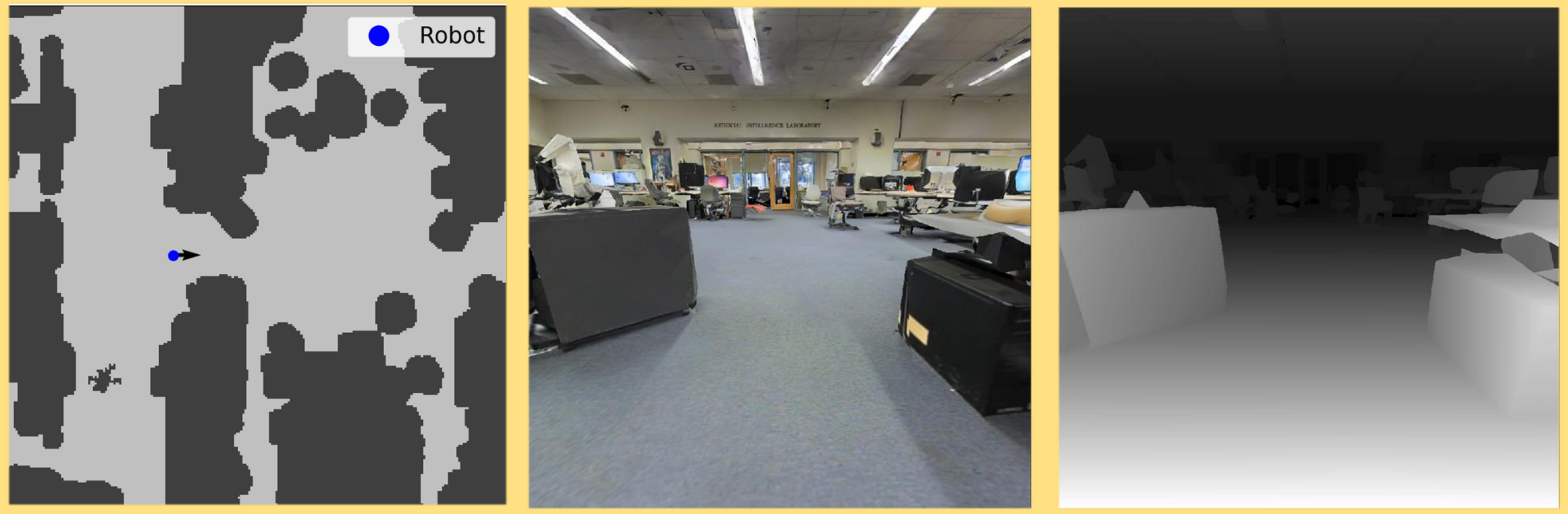
Human State & Control

$[x, y, \theta, v, \omega]$

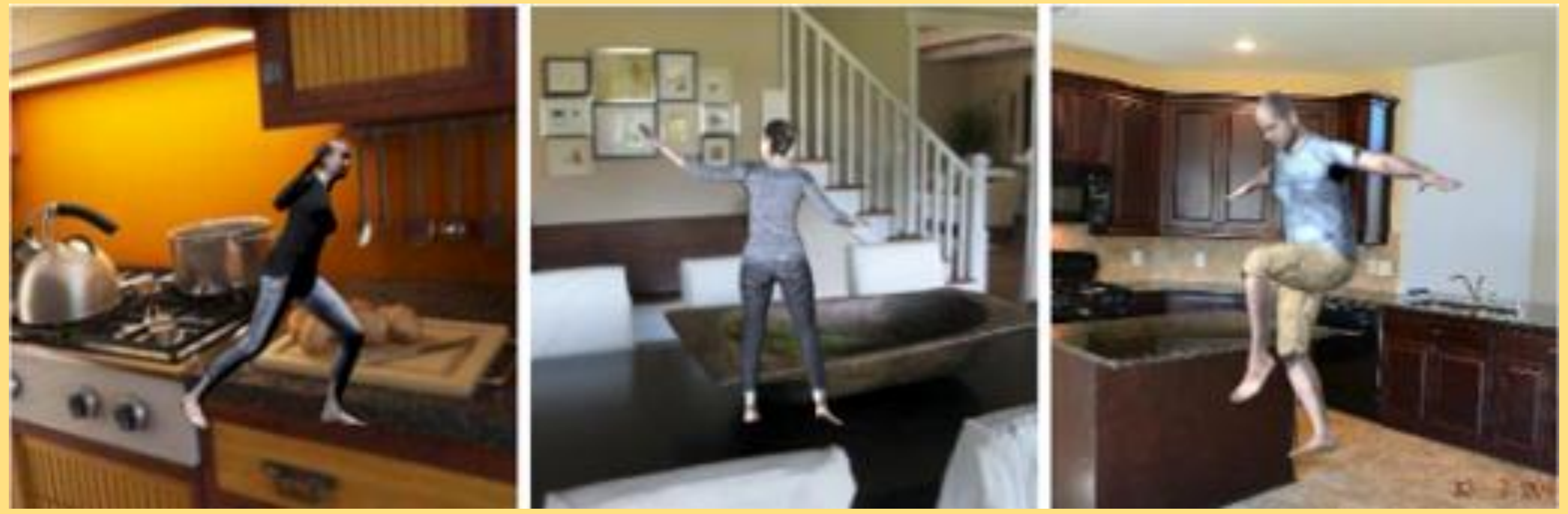
Human Identity

- Gender
- Texture (Clothing, Skin Color, Facial Features)
- Body Shape (Tall, Short, etc.)

SD3DIS



SURREAL



HumANav

Robot State

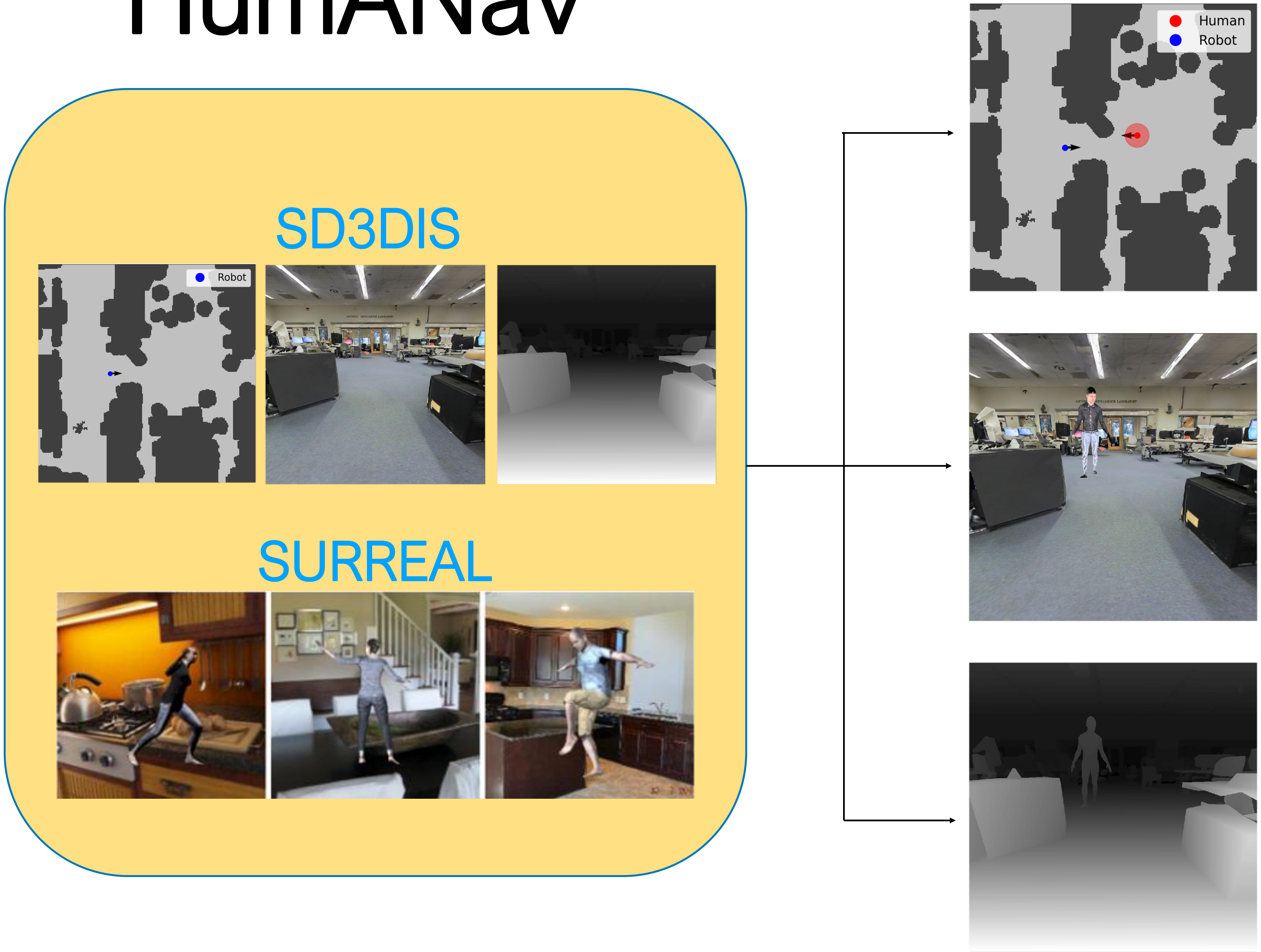
$[x, y, \theta]$

Human State & Control

$[x, y, \theta, v, \omega]$

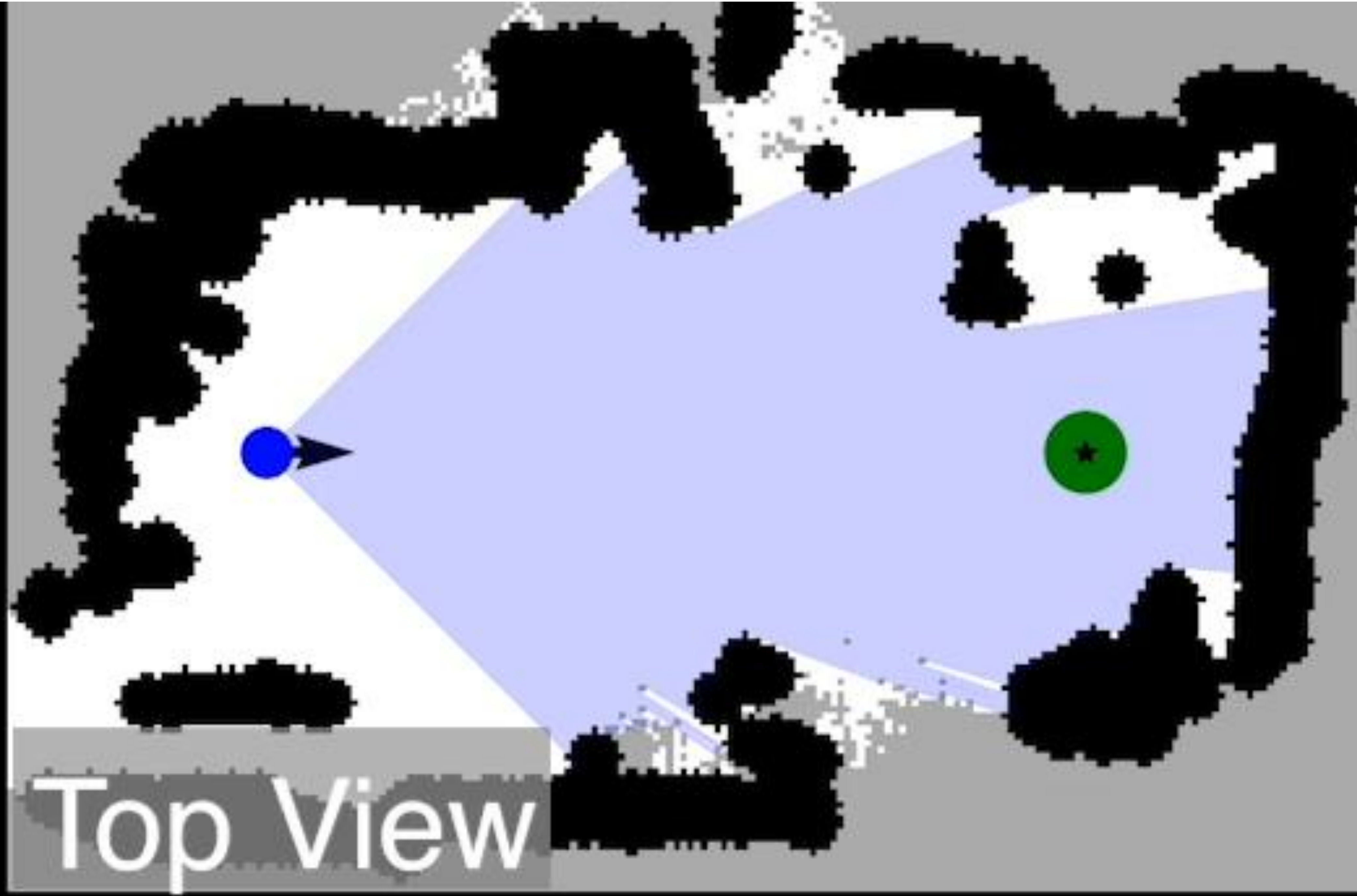
Human Identity

- Gender
- Texture (Clothing, Skin Color, Facial Features)
- Body Shape (Tall, Short, etc.)



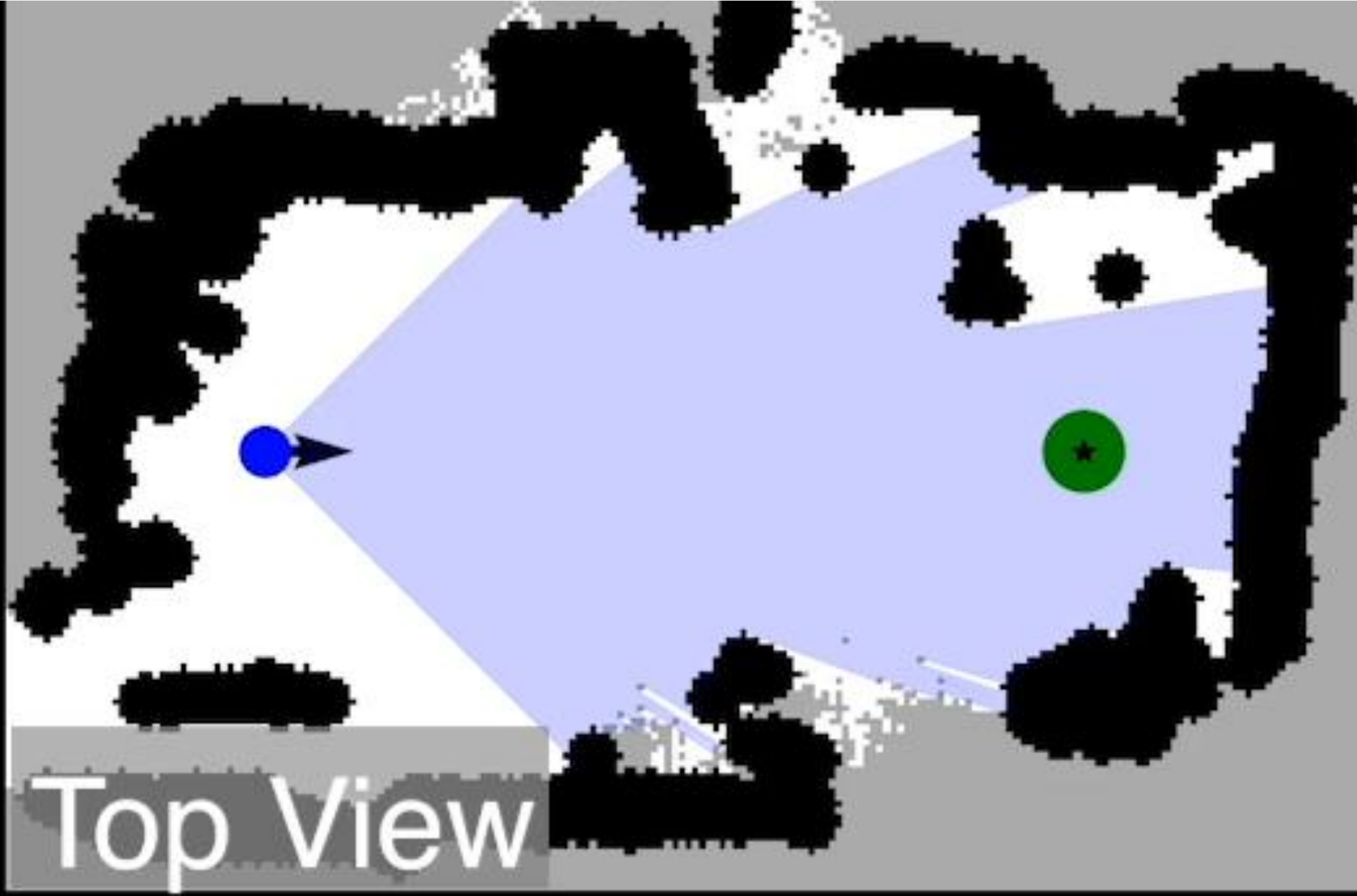
LB-WayPtNav

Experiment 1 (1x)



LB-WayPtNav-DH

Experiment 1 (1x)



Summary

Incorporating perception in the control loop

Supervising learning using optimal control

- a perception-planning-control pipeline
- comparison with a more traditional SLAM pipeline
- applied to a vision-based navigation task

Models of human motion

Challenges for safe control