



Toward Human-Like Social Robot Navigation: A Large-Scale, Multi-Modal, Social Human Navigation Dataset

Duc M. Nguyen, Mohammad Nazeri, Amirreza Payandeh, Aniket Datar and Xuesu Xiao

Proposed Work:

An *open source*, *first-person-view*, social human navigation data collection sensor suite.

A Multi-modal Social Human Navigation Dataset (MuSoHu): a large-scale, egocentric, multi-modal, and context-aware dataset of human demonstrations of social navigation.

Analysis in terms of human and robot social navigation and point out future research directions and anticipated use cases of our dataset.

Social Navigation for Mobile Robots

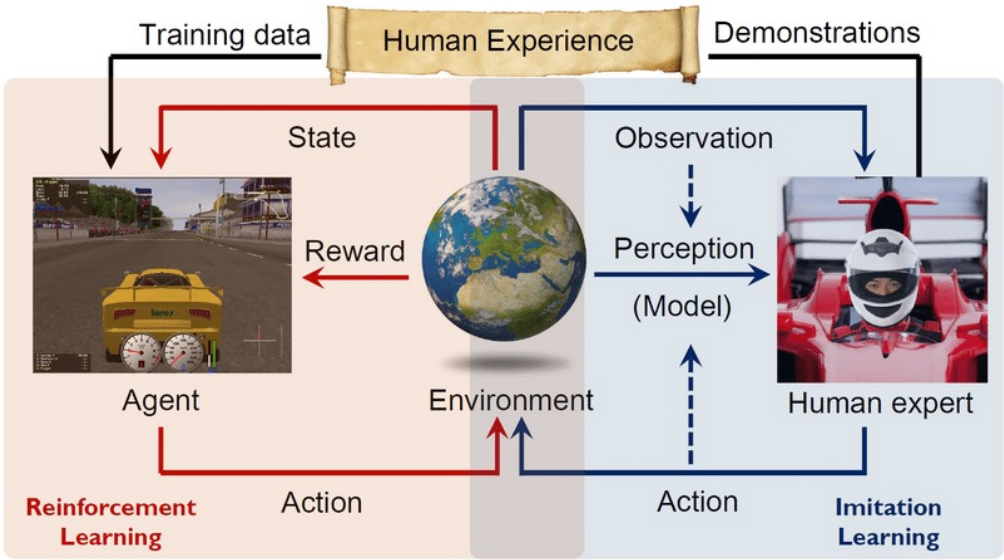
- Socially aware robots are being used in various applications, but their ability to work among humans needs to be improved.

- Classical planners often struggle to be socially compliant.

Current Approaches

- Reinforcement Learning: learning from trial-and-error experience.
  - RL in the real world is extremely expensive due to the limited availability of robots.
  - RL in simulation requires a good model of social navigation interactions of humans.
- Imitation Learning: learning from expert demonstrations.
  - IL requires demonstration datasets collected on robot platforms, mostly through expensive human teleoperation at scale.

=> Both practices need large datasets.



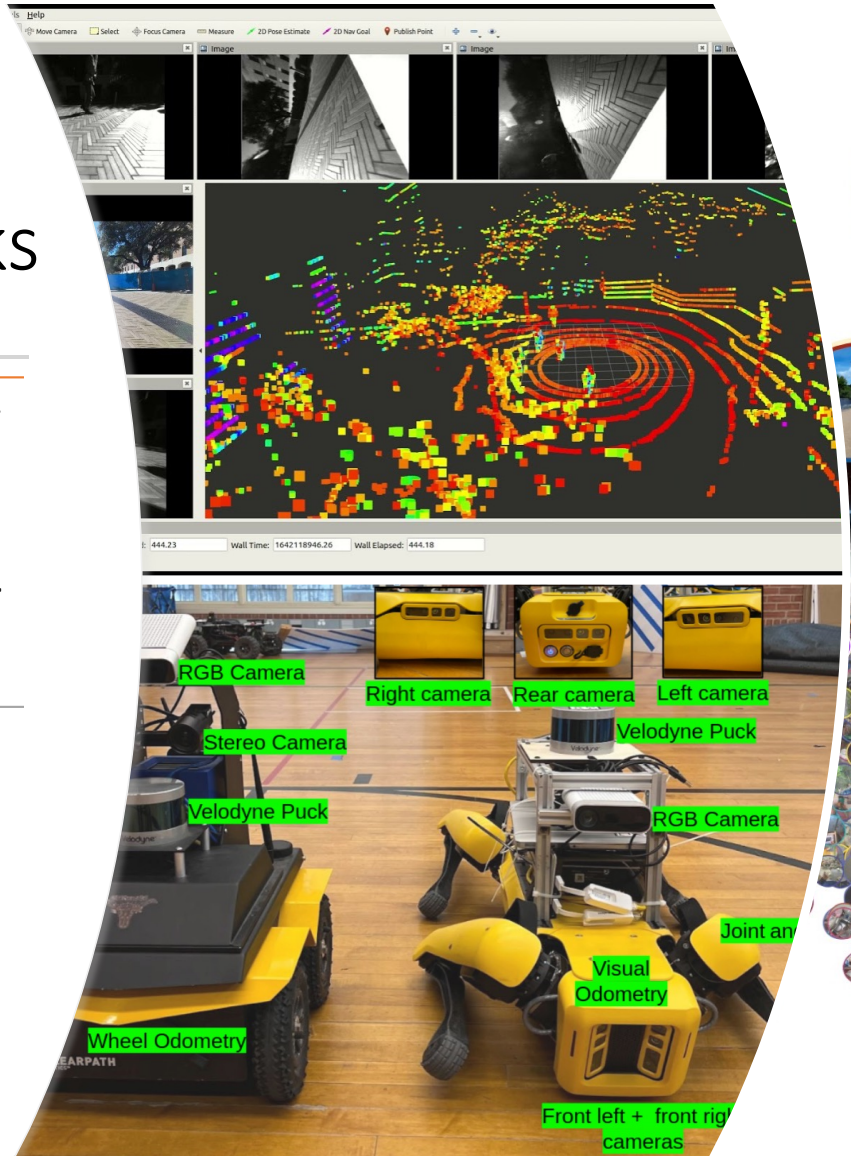
Zuo, Sixiang & Wang, Zhiyang & Zhu, Xiaorui & Ou, Yongsheng. (2017). Continuous Reinforcement Learning From Human Demonstrations With Integrated Experience Replay for Autonomous Driving. 10.1109/ROBIO.2017.8324787.

Related Works

Ego4D: 3,670 hours of daily-life activity video spanning hundreds of scenarios captured by ***931 unique camera wearers*** worldwide.

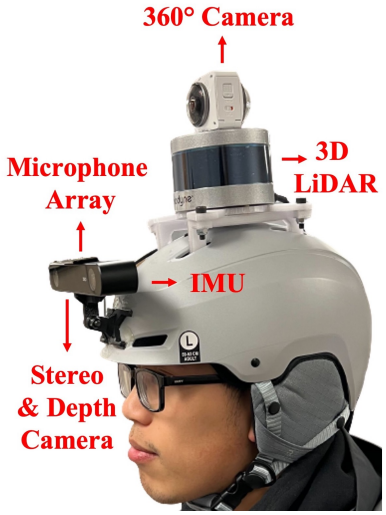
[K. Grauman *et al*.]

SCAND: 25 miles and 8.7 hours of ***robot driven trajectories*** through a variety of social environments around the University of Texas at Austin campus. [H. Karnan *et al*.]



Sensor suite

- Designed to gather navigation data from ***human movements*** captured by ***robotic sensors***.
- The ROS program used to control the sensors is published on GitHub.
- Sound Information:
  - Microphone Array (Seeed Studio ReSpeaker Mic Array v2.0)
- Spatial Information:
  - 3D LiDAR (Velodyne Puck)
- Visual information:
  - Stereo & Depth Camera (Stereolabs ZED 2)
  - Overhead 360 Camera (Kodak Pixpro Orbit360 4K)

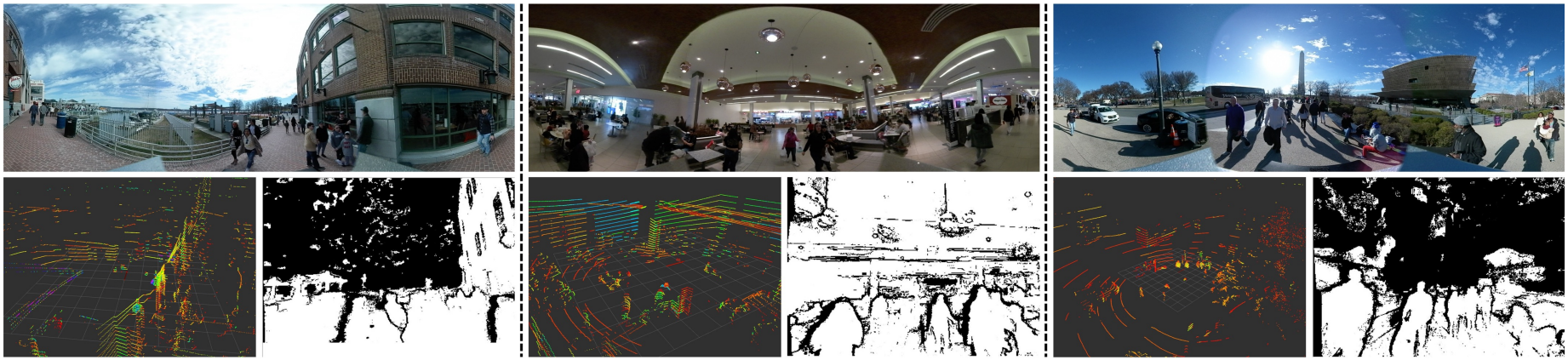


ROS

Logo of ROS Noetic Ninjemys

MuSoHu Dataset

- Our dataset contains *~**100** km, **20** hours, **300** trials, **13** human demonstrators*.
- The portability of our sensor suite allowed the dataset to be collected in different places:
  - Virginia: GMU campus, main streets (Fairfax), Old Town, Alexandria, Springfield Towncenter.
  - Washington DC: Georgetown Area and 18th St NW.
  - Maryland: Tanger Outlets National Harbour.



MuSoHu Dataset

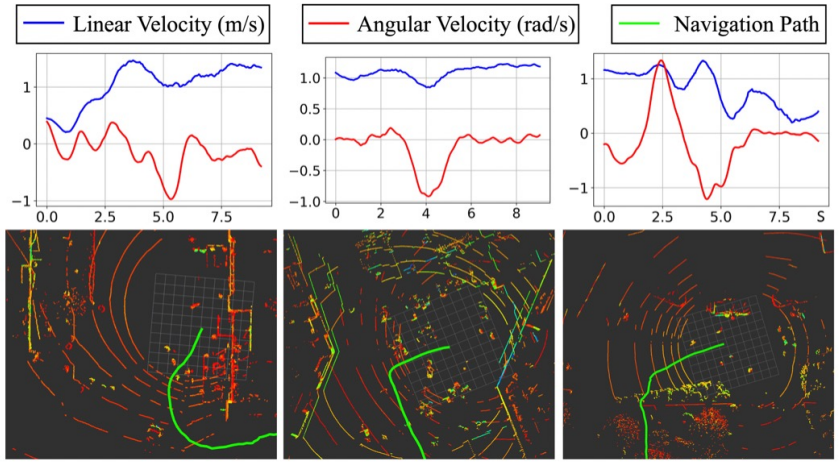
TABLE I: Descriptions of Label Tags Contained in MuSoHu.

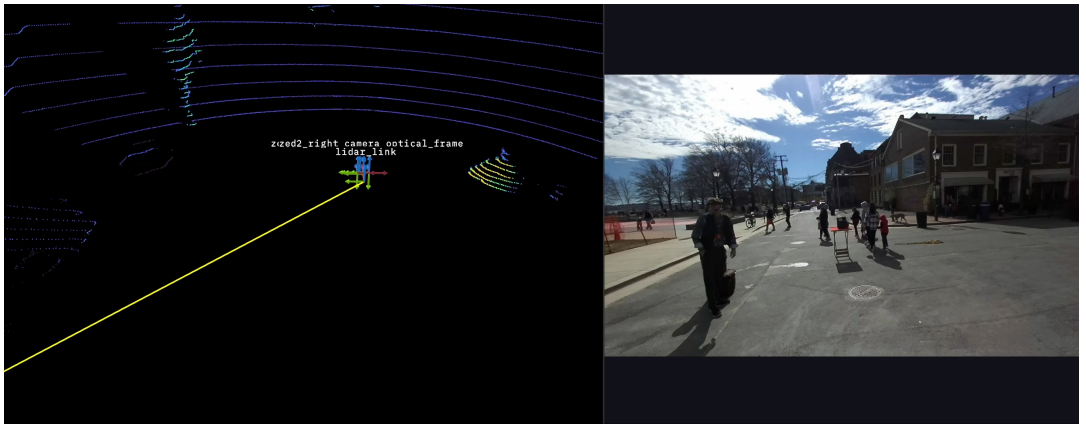
Tag	Description	# Tags
Against Traffic	Navigating against oncoming traffic	124
With Traffic	Navigating with oncoming traffic	90
Street Crossing	Crossing across a street	51
Overtaking	Overtaking a person or groups of people	62
Sidewalk	Navigating on a sidewalk	60
Passing Conversational Groups	Navigating past a group of 2 or more people that are talking amongst themselves	52
Blind Corner	Navigating past a corner where the human cannot see the other side	46
Narrow Doorway	Navigating through a doorway where the human opens or waits for others to open the door	23
Crossing Stationary Queue	Walking across a line of people	24
Stairs	Walking up and/or down stairs	17
Vehicle Interaction	Navigating around a vehicle	13
Navigating Through Large Crowds	Navigating among large unstructured crowds	19
Elevator Ride	Navigating to, waiting inside, and exiting an elevator	7
Escalator Ride	Navigating to and riding an escalator	2
Waiting in Line	Waiting in Line to enter congested areas	2
Time: Day	Navigation during day time	65
Time: Night	Navigation during night time	32

- It is very easy to collect data at anytime, and anywhere.
- Each demonstration goes with a list of social interactions and contexts so researchers can easily study context-aware social navigation or certain behaviors from MuSoHu.

Analysis

- Policies can be learned from MuSoHu dataset as datafiles provide twist messages (/action) besides the synchronous data streams from sensors.
- The following figure shows corresponding linear and angular velocities (filtered by Savitzky-Golay filter to smooth out high frequency noises caused by walking gait) and navigation path taken by the human demonstrator in three scenarios.





Analysis

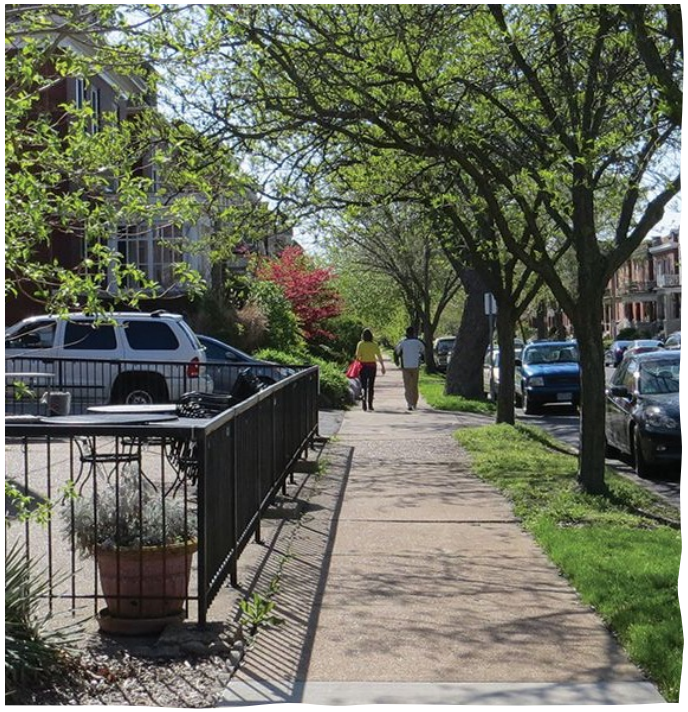
- To emphasize the utility of MuSoHu dataset, we trained a navigation policy, using Behavior Cloning technique, for two different robot platforms (Hunter SE and Go1 robot dog) to handle path planning when approaching humans.

Future use cases

- Study social human v robot navigation:
  - Compare humans' ways of navigation to a robot and evaluate what really makes humans so good at this task.
- Study how humans signal our navigation intention:
  - Maybe robot should also be able to "gaze" or use gestures.
- Real2Sim Transfer for Social Navigation:
  - Researchers can use MuSoHu for synthesizing more interactions to create a more detailed simulator.

(a) Looking right (b) Looking straight (c) Looking left

(d) Indicating right (e) No indicating (f) Indicating lef



Future use cases

- Learning social robot navigation:
  - Our sensor suite is easy to replicate, open the door to do culturally dependent social navigation as data can be collected from different countries with different social “norms”.
- Imitation Learning from mismatched observation:
  - While it’s true that a policy can be derived from humans’ demonstrations, a robot’s set of sensors can be positioned very differently (most robots are lower than humans)





Toward Human-Like Social Robot Navigation: A Large-Scale, Multi-Modal, Social Human Navigation Dataset

Duc M. Nguyen, Mohammad Nazeri, Amirreza Payandeh, Aniket Datar and Xuesu Xiao