GO, LOOK, AND TELL: CONVERSATIONAL, DYNAMICALLY STABLE MOBILE ROBOTS

Matthew Wilson, Jean Oh, Ralph Hollis

Carnegie Mellon University

MOTIVATION

Robots will become much more useful when they can effectively communicate with humans and better understand the world

- More natural interactions enable new uses for robots
- Service and care industries. Household personal robot?

OBJECTIVE

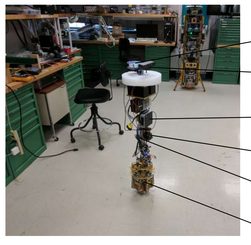
Develop a system to allow the robot to be commanded by speech and respond to user questions about the environment

- Baseline integration of components: navigation, perception, and speech
- Extensible to more sophisticated natural language understanding and probabilistic world modeling

To demonstrate the system, be able to command robot to *go* to the office, *look* what is on the table, come back, and *tell* us what it saw.

BALLBOTS

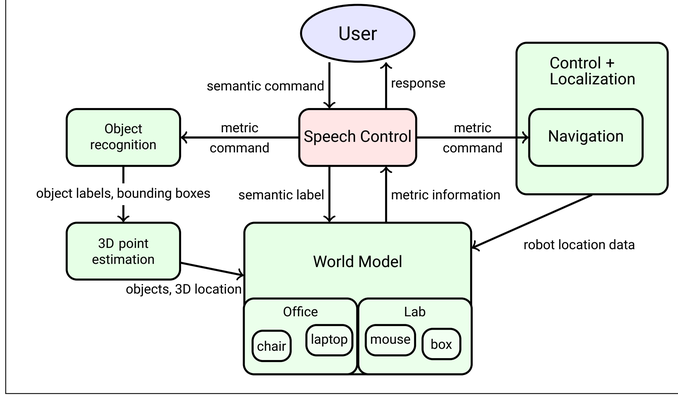
Robots that actively balance on a ball using an inverse mouseball drive (IMB) mechanism. They are omnidirectionally compliant and well suited for human robot interaction. This one is called Shmoobot, or Shmoo.



Cool things:

- Amazon Echo Dot for speech control
- Astra RGB-D camera for object detection
- Intel NUC, Ubuntu 14.04 w/ ROS Indigo
- Hokuyo laser for navigation
- VectorNav IMU for sensing lean angle
- Inverse mouseball drive mechanism

APPROACH



DEMONSTRATION

"Go to the office. Look what is on the table. Come back to the lab. Tell me what you saw."

"Got it"

Plans a path and autonomously navigates [1]-[2] to the office

In the office, turns to look at the table and see what is there

Detects objects on table using YOLO object detector [3]

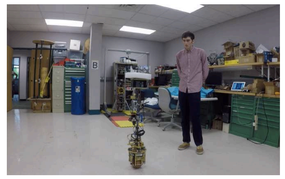
Estimates 3D location of objects by segmenting point cloud data based on bounding boxes

"What have you seen?"

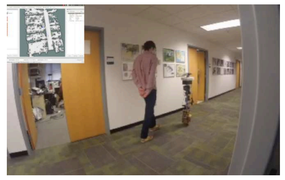
"I have seen three monitors, a mouse, a chair, a keyboard, a cup, a bottle, and a laptop"

Displays images and 3D locations of detected objects on map

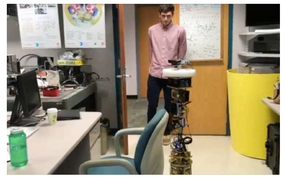
Speech command to control robot



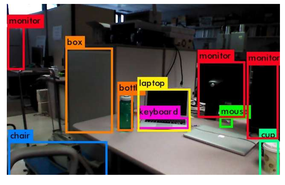
Path planning and navigation



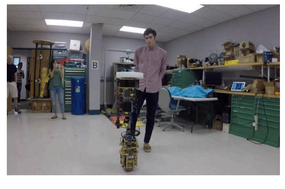
Observing table



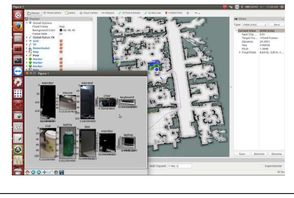
Object detection and localization



Ask robot what it saw



Objects and 3D positions on map



CONCLUSION

- Integrated navigation, perception, and speech control on a dynamically stable mobile robot to create a full, capable system
- Allowed user to command robot to perform dynamic tasks interactively and to communicate world knowledge
- System modularity will allow replacement of components with more intelligent and adaptive ones

FUTURE WORK

More sophisticated natural language communication

- Probabilistic grounding of objects referenced in commands to real world objects
- Representing spatial relations (ex. to the left of the box)

Multi-hypothesis world model

- Maintain multiple beliefs of world, based on uncertainties
- Keep track of objects over time

More interactive commands and responses

- Clarification for ambiguous goals
- Confirmation of actions
- Update belief of objects based on corrections given by user

REFERENCES

[1] M. Shomin and R. Hollis, "Differentially flat trajectory generation for a dynamically stable mobile robot," in *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, 2013.

[2] M. Shomin and R. Hollis, "Fast, dynamic trajectory planning for a dynamically stable mobile robot," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2014.

[3] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger", arXiv preprint arXiv:1612.08242, 206

ACKNOWLEDGEMENTS

- Thank you to Roberto Shu  for help in conducting experiments
- Thank you to the NSF for funding this research