Integrating Action Response In RAS

Asim Fauzi[2], Glenn Duncan[1], Dr. Larry Holder[1], Dr. Aaron Crandall[1]

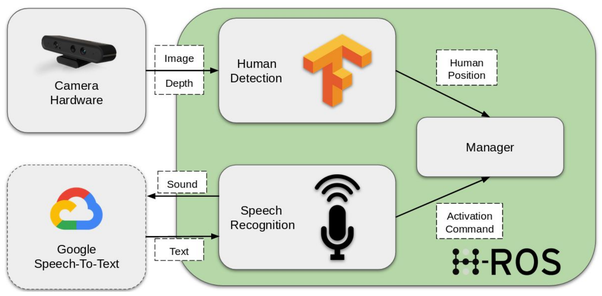
Washington State University[1], Voiland College of Engineering and Architecture
Miami University[2], College of Engineering & Computing

Background

- Robot assisted living could be crucial for the future of healthcare as humans are living longer than before and an inflated elderly population has induced an increased demand in eldercare labor.
- Robot assisted living is not yet prevalent in the modern day healthcare industry.
- Robot Activity Support, or RAS, is a prototyped assisted living robot whose main purpose is to aid Alzheimer's patients and other cognitively impaired individuals.
- In previous research, RAS could approach you when an error was made in an activity.[1]
- The current objective is to implement a response system for RAS where a patient may call out to request help and receive a proper response such as approaching the human and prompting assistance.

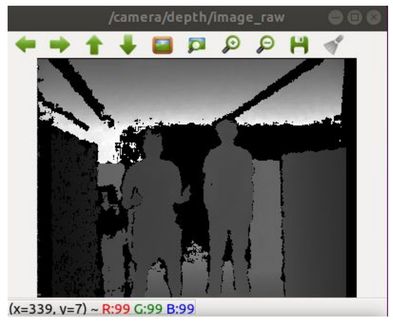
Approach

- RAS is built using Robot Operating System (ROS) which comprises of a master nameservice that communicates with multiple processes[2].
- The speech recognition node uses Google's *Cloud Speech-to-Text API* to recognize key trigger words which initiates speech-to-text transcription.
- The human detection node uses Tensorflow's *faster rcnn inception* model.
- Once the human requester is found, RAS calculates their locally mapped position on a 2D grid.



Methods & Procedures

- This experiment involved a number of participants aged 18+ to perform a scripted multi-step task.
- The participants were asked to call RAS over, select a choice from RAS for assistance (e.g. play entire video, play step video, etc.) and then dismiss RAS.
- The process was repeated 3 times for each participant, with each repetition having a different end point for the robot to navigate to.
- RAS was rated on how well it performed its task
- Performance metrics were end distance from the target user, activation and command understanding accuracy, and human detection accuracy.



Results

- Navigation Performance
  - Average distance from participant: ~5 feet from the user
  - The robot typically performed best at navigation when the user was farther away from it to start.
- Human detection
  - 91% Accurate in finding humans
- Speech Recognition
  - Activation Accuracy: 62.6%
  - Command Accuracy: 85.7%

Analysis & Conclusion

- Design Choices for the speech interface
  - Maybe "Jasper" isn't the best keyword.
    - Many speech recognition systems use a keyword that's easier to say, such as one that starts or ends on a vowel, or one with multiple syllables.
  - Making the system more event-driven would increase performance in the speech recognition.
  - Currently, the system doesn't use ambient volume, so that is likely beneficial to make the system easier to interact with.
- Human Detection and Navigation
  - It would be beneficial to incorporate further object detection capabilities into RAS (e.g. recognizing tables to avoid them.)
  - It would also be beneficial to ensure that RAS faces the human it's interacting with.

Future Work

- One future application of this work could be incorporating emergency response to RAS, by allowing it to locate and navigate to a user outside of its field of vision.

- Another application can be used for emotion recognition. The current human detection setup would then be able to monitor emotions of the humans it interacts with which can be valuable data for predicting mental and physical health.

References & Acknowledgements

[1] Wilson, G., Pereyda, C., Raghunath, N., Cruz, G. D., Goel, S., Nesaei, S., . . . Cook, D. J. (2019). Robot-enabled support of daily activities in smart home environments. *Cognitive Systems Research, 54*, 258-272. doi:10.1016/j.cogsys.2018.10.032

[2] V. M. (2014, June 15). Wiki. Retrieved May 27, 2019, from http://wiki.ros.org/ROS/Technical Overview

This project was funded by NIH grant 1R25AG046114 and NSF grant #1734558.