Improving HRI Capabilities for Serving Robots by Combining Visual and Aural Cues[1]

Hari Krishna Prannoy Namala

[a] *University of Maryland, College Park, 20740, Maryland, USA*

Human body's sensing capabilities allow us to interact with the world and other humans. Similar to humans, an autonomous robot interacting with humans on a regular basis needs to have well equipped sensor systems. Applications of robots with such capabilities are serving at events, waiting tables at restaurants, elderly care where robots can be helpful working alongside humans enabling humans to use the saved time to perform other tasks and interact with humans at levels in which a robot cannot. The capabilities of humans to work in noisy environments has always been a topic of intrigue for autonomous robot researchers. Consider an example of waiters at restaurants, the noise levels at restaurants are high. Human waiters are able to navigate such environments without much difficulty most of the time. Achieving this level of autonomy is a really hard task in a robot. However, with the recent advancements in the sensory systems for robots, it is possible to integrate these senses to create a robot which can deal with lower levels of noise in the environment.

In this study, we explored the capabilities of current state of the art for sensor systems in autonomous robots. Based on the analysis, an approach is proposed for a lightweight audio and video based human tracking system for service robots. The problem is split into three components: Sound Source Localization - to locate the customer in need, Wakeword Detection - as an alert to look for the customer and Human Body Posture Detection - to identify the customer in a group. In the following paragraphs hardware and software used to implement the subsytems and the techniques used to integrate them will be summarized.

[1] This is summary of the ENPM808: Independent Study under Dr. Dinesh Manocha. Github link with the code contribution - here

Firstly talking about Sound Source Localization. The technique of sound source localization used for this study is General Cross-Correlation with Phase Transform (GCC-PHAT). One of the implementations of GCC-PHAT is Open Embedded Audition System (ODAS). ODAS is a framework providing artificial audition capabilities in real-time while running on low-cost hardware. To run ODAS, we used MATRIX Creator. MATRIX Creator is a microphone array with 8 microphones arranged in a circular configuration. The microphone array is attached to the GPIO pins of a Raspberry Pi 3B with Raspbian Buster OS. The sound localization program with ODAS is run continuously to give DOA of any sounds the microphone is hearing.

For Wakeword Detection, we use Porcupine Wakeword Engine from PicoVoice. Porcupine is a highly-accurate and lightweight wake word engine. It enables building always-listening voice-enabled applications. It can also be run on the Raspberry Pi alongside the sound source localization. The output from the wakeword detection will be a prompt whenever the word is detected. Once the wakeword is detected, the output from the localization program will be recorded. Once the angle is recorded, it is used as a reference by the navigation stack of the robot to turn towards the output angle.

Lastly, for Human Body Posture Detection, we use the pretrained models available on the Zed 2i camera. The camera comes with the SDK which has the models for human body detection. The model describes human body in 18 points. Using this data, a simple if-else logic is used to recognize a hand raise by tracking the points which represent the hand joints.

The system integration framework used is ROS. ROS has a basic communications version which can be used on a Raspberry Pi. The process flowchart is shown in figure.

The individual components of the system were working perfectly. A standalone sensor system was designed which was tested by placing the system on a static mount. The integrated system works as expected. Due to time constraints, the integrated system was not coupled with the navigation stack of the robot.

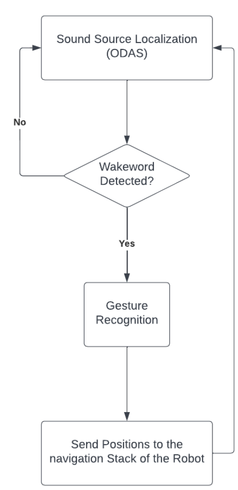

Figure 1: Sensor System Process Flowchart