Pepper

The Artificially intelligent personal assistant robot.

Pepper is a differential drive robot that is capable of mapping an environment and navigating in it and is aimed at being a personal assistant for the user using various artificial intelligent algorithms. Few of the algorithms used can perform - speaker recognition, object detection and recognition etc. A gripper is also present on the chassis of the robot that performs object grasping. The robot uses a ROS network as a platform of communication between the various sensors and actuators and the on board computer (ROS master).

**Robot Design:**

The chassis of the robot was designed using Autocad and the design was fabricated using acrylic. The chassis can be broadly classified with three layers - bottom layer to contain the battery and driver, middle layer to contain the gripper arm and the top layer that contains the on board computer. Microsoft Kinect is attached to the topmost pedestal. Two high torque motors with internal rotary encoders are used to move the robot and a castor wheel is used to provide stability.

**Mapping:**

To navigate an environment optimally its map is required. Pepper achieves this using the on board kinect sensor and the gmapping ROS package. The package uses Simultaneous Localisation and Mapping Algorithm (SLAM) to generate the map.

**Localisation:**

As the size of the robot increases the drift in the wheels also increases. To tackle this problem we use more than one sensor and perform Kalman filter algorithm. The inbuilt encoder is used to perform dead reckoning and obtain the pose of the robot. A gyroscope is also used to obtain the yaw of the robot and these sensor values are used in the Kalman filter to estimate the pose.

**Speaker Recognition:**

When the robot detects a voice command from a user, it must first identify the speaker before responding to him. A machine learning algorithm was developed to identify the speaker of a voice sample. The algorithm used was text independent and developed as a real time application. The voice samples of the speaker were first trained using the Mel Frequency Cepstral Coefficient(MFCC) Vectors, extracted from the training voice data. The voice model of the speaker was thus developed and the corresponding parameters saved in the database. This was done for all speakers. When a real time test sample is detected, the machine extracts the MFCC vectors and compares it with the already saved voice samples. The closest match in the existing database is found and hence the speaker is recognised.

**Object Recognition and Pick up:**

The bot is trained with 1000 classes of object using Convolutional Neural Network (CNN). CNN is a bio inspired machine learning algorithm and its functions are similar to the eye. The input image is captured using kinect and its depth is measured. There is a robotic arm mounted on the base of the robot. If the depth is greater than the workspace of the robotic arm then the robot moves closer to the object. The robotic arm has 3 degrees of freedom (gripper has 1 DOF). The object is recognised and the robot picks up the object.