3D Camera Localization - Probabilistic Robotics

Francesco Chichi

September 12, 2017

1 Introduction

This project is a simulation of 2D bearing only SLAM, i.e., Simultaneous Localization and Mapping, for a differential drive robot called "Orazio". Orazio is controlled by translational and rotational velocities inputs and its location lies within a 2D map. The robot can detect a set of uniquely distinguishable landmarks through a 2D landmark bearing sensors in order to reconstruct the position of all the observed landmarks in the map (in SLAM problem the location of the landmarks in the world is not known). The software for the project was implemented in C++ and the input datasets are imported from G2o, i.e., a General Framework for Graph Optimization.

2 Data Structure

2.1 Basic Autoencoder

3 Least-Squares Approach

3.1 Prediction

In order to build our neural network we used two set: the training set extracted from the CHiME Dataset, a dataset containing several hours of recording of living room's sound where some people perform common actions such as talking, playing; the testing set, the same used by paper's author, where the audio files were augmented with abnormal sounds like screams, alarms, etc in order to simulate novelty sounds and test the novelty detection. Both the sets are built in the same way, except for the storage of labels that are extracted only for testing set. The building of sets can be split in two parts: data extraction and building of the set.

3.2 Correction

In order to recognize an input as novelty sound the network uses an adaptive threshold in the testing process, in particular in the *test* function. The idea behind the threshold is that we trained our neural network in order to reconstruct each normal sound, so if it will receives a normal sound as input, the reconstruction error should be low, while if it is very different from the inputs that neural network saw in training phase, therefore if it is a novelty sound, the reconstruction error should be greater.

Following the paper's approach we calculate for each sound the euclidean distance, on a frame basis, from the input value and the reconstruction error, then we calculate the mean of the relative frame. Finally we calculate the median of the means for each audio.

4   Experimental Results

5   Front-End