RGB-D Vegetable Recognition

Introduction

RGB-D signifies the Red, Green, Blue and Depth data of the percept obtained using a Microsoft Kinect® style 3D sensor. The data obtained using the sensor comprises of data points of the order of 100K, each point describing the color and the depth of each pixel as observed from the sensor. The widespread, and low cost availability of these sensors and hence the data has enabled the contemporary research methods to extract useful information from the percept. Especially in applications involving autonomous navigation, semantic inference, object recognition etc. where a computational algorithm derives essential information from its surrounding environment, the use of RGB-D data has found major interest. Mainly due to the additional depth cue that the data provides, researchers from the image processing and computer vision domain have incorporated the depth information to further optimize the performance of the existing RGB based algorithms.

Exploiting these factors of the RGB-D dataset, OwlParrot's Future Fridge project is investigating the use of 3D recognition framework for detection of the type and location of vegetables present in the refrigerator. Future Fridge is an autonomous refrigerator embedded with state of the art computational technologies, and cutting edge hardware to service the extraction and cutting of the vegetables in accordance with the requirements of the user. The autonomous agent is capable of adapting and learning the user preferences and mine additional information from fellow instances of Future Fridges as well as the world wide web. Providing its users with remote connectivity and an easy to use mobile interface, the refrigerator plays the role of a assistant chef by keeping the vegetables and edibles ready even before the user returns home.

RGB-D recognition serves as the sole perception and planning module for the execution of the cutting tasks. Depending on the requirements of the user, this module extracts information about the available vegetables and their 6D pose information relative to the racks of the refrigerator. This forms the main input of the motion planning module of the robotic manipulator, which then moves the manipulator arm to extract the required vegetables and forward them to the cutter hardware. This document details out the process involved in RGB-D vegetable detection.

Requirements

Vegetable recognition framework involves intricate steps in order to imitate the perception capacity of humans. To maximize the performance and accuracy of recognition, various factors need to be addressed. Some of key implementation hurdles include the following:

- **Pre-processing:** to ensure the obtained RGB-D data is free from measurement noise
- **Vegetable Detection:** The racks in the fridge contain either planar surface or small sections where the vegetables are placed. The additional surface pertaining to parts of fridge perceived are removed using appropriate segmentation algorithms
- **Feature Extraction:** The residual clod is processed to extract identifiable features from the vegetables present in the point cloud.
- **Recognition:** Finally, the features are matched with stored repository/web to identify the vegetable type, its size and location.

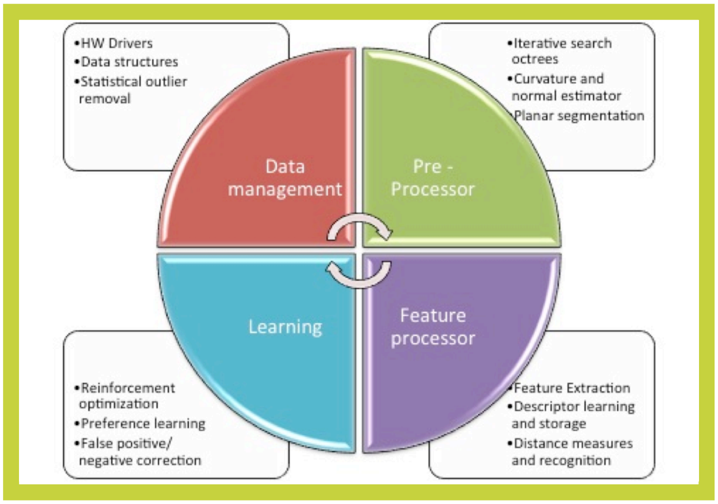

Figure 1: Overview of the underlying software framework

Framework

The capacity of handling huge amount of data requires well-defined data structures and robust processing backend. To this end, the current framework comprises of constructs capable of handling large 3d-point clouds and a minimal set of computational algorithms for processing them. The overall software framework underlying the recognition module is shown in Figure 1. Major parts include the (a) Data management sub-module, (b) Pre processing sub-module, (c) Feature processing sub-module, (d) Learning sub-module. This document futher describes the details about the roles, and working of each of the sub-modules.

Data Management

Each point in the obtained input comprise four floating-point numbers viz. R, G, B and the depth information. Inefficient use and iterations performed on large quantity of such points can lead to computational complexity. Therefore, use of flexible data structures for their storage is mandatory. This sub module, apart from performing the role of communication with the hardware sensors, provides underlying data storage, retrieval and filtering constructs. The module provides various options for storage of the raw data are stored in the memory. Depending on the type of operation being performed, a choice between intensity image, RGB image, organized depth image, or even registered RGB-D point cloud can be obtained. Other functionalities include easy data manipulation viz. concatenation, deletion, indexing, implementation of iterative constructs, memory management and clean up. The outlier removal technique employs statistical techniques to compute the density of points in a given location of the cloud, and filters the data based on given density threshold, thereby removing points added into the cloud due to sensor noise.

Pre Processor

To extract relevant features from the raw data, essential summaries pertaining to small subsets of points are required. This module extracts normals of a small part of 3D surface, geodesic distances, convex hull of a subset etc. Further, for fast iteration through all points, this modules also implements an Octree search algorithm.

Feature Processor

Recognition in 3D domain works on the principal of categorical identification of pre-learned features pertaining to different types of vegetables. This module employs state-of-the-art features and descriptors for the identification process. The module is first pre-trained to recognize a wide range of vegetables. This is done by exposing the feature-learning algorithm with different views, sizes and colors of each vegetable one at a time. The algorithm is informed about the vegetable being exposed. By this process, the algorithm develops a collection of feature descriptors for each vegetable type. To ensure high accuracy, the feature extraction process is implemented with view and scale-invariant feature descriptors such is SIFT or SURF. Further, a bag-of-words learning and recognition framework is used to match the extracted features from the ones existing in pre-learned dataset. The pre-trained algorithm uses a hierarchical classification system based on colors of the vegetable. The hierarchical process first considers the color, and then performs feature matching only on the descriptors that match the color. The algorithm also provides extension points for ease of integration with additional image based classification system.

Learning

Reinforcement forms a major part of the vegetable recognition framework. Different parts of world may contain different varieties of the shape/size/color of a same vegetable. However, training the fridge to recognize all of them out of box is impossible. Therefore, an additional reinforcement algorithm implemented in this framework learns by experience. The user can customize the fridge to correct its beliefs about a type of vegetable by providing reward on successful recognition, and a penalty on negative recognition. Apart from learning the vegetable class, this sub-module also records user preferences about locations, color and shapes of vegetables found in the fridge and updates its descriptor database for increased accuracy.