Chapter 21
Object Recognition Pipeline:
Grasping in Domestic Environments

John Alejandro Castro Vargas
University of Alicante, Spain

Alberto Garcia Garcia
University of Alicante, Spain

Sergiu Oprea
University of Alicante, Spain

Sergio Orts Escolano
University of Alicante, Spain

Jose Garcia Rodriguez
University of Alicante, Spain

ABSTRACT

Object grasping in domestic environments using social robots has an enormous potential to help dependent people with a certain degree of disability. In this chapter, the authors make use of the well-known Pepper social robot to carry out such task. They provide an integrated solution using ROS to recognize and grasp simple objects. That system was deployed on an accelerator platform (Jetson TX1) to be able to perform object recognition in real time using RGB-D sensors attached to the robot. By using the system, the authors prove that the Pepper robot shows a great potential for such domestic assistance tasks.

INTRODUCTION

Nowadays, social assistant robots have become a way of improving people's quality of life in modern societies by performing domestic tasks. Life expectancy is in continuous growth, causing an increase in the average age of the population. For this reason, the assistance of elderly people becomes a priority for developed countries, which consider social robot assistance as a good initiative for helping dependent persons. This technological approach provides an alternative to the use of qualified human personnel, reducing economic costs in the long term.

DOI: 10.4018/978-1-5225-8060-7.ch021

A social assistance robot could perform different tasks such as object detection (identify a particular object from the scene), object grasping (with the purpose of bringing objects to the user), interaction with the user by means of gesture and speech recognition, and programmed actions like home cleaning. However, it is very common to design different robots for each particular task instead of integrating all tasks in the same one. Therefore, we consider that a robot equipped with arms and hands would be capable of performing both recognition and grasping tasks. Following this idea, in this project we will perform object recognition alongside grasping tasks in unstructured environments. To that end, we will leverage a humanoid robot such as Pepper designed for social assistance.

Manipulating an object without knowing its position and pose is not straightforward and implies the use of techniques from different areas. On the one hand, an object recognition process is needed in order to detect the objects in the scene. That system must provide information about object positions and orientations. This problem has been addressed in several works, mainly using RGB cameras alongside depth information in order to obtain a partial reconstruction of the environment which can be used to recognize objects. On the other hand, once positions and orientations are known, the hands and arms of the robot have to be moved with the purpose of grabbing the target object. This process is not straightforward since many problems regarding both the object (different texture, size, weight, or shape) and the variation of robot arms and hands degrees of freedom must be solved.

BACKGROUND

Object recognition is the process by which objects are detected in images, obtaining information related to their position and orientation in the scene (Garcia-Garcia et al., 2016). There are several approaches for this purpose but we will emphasize those ones which are based on local features since they are more robust in unstructured environments with occlusions. These techniques extract representative local features from both the scene and the models and then allowing identify those model objects by matching the extracted features.

The classical pipeline for object recognition based on local features is three-staged (Guo et al., 2014). First, several keypoints are detected in order to extract representative information from the scene. This will improve the computational cost of the pipeline by processing and discarding ambiguous regions that do not provide important information. Next, the neighborhoods of those keypoints are described for the training or matching stage by encoding them into descriptors (Bronstein et al., 2010). Finally, correspondences between descriptors from the scene and the model objects are obtained. This last stage of the recognition pipeline can be further divided into three steps (Guo et al., 2014): (1) matching or correspondence search, frequently using techniques like Nearest Neighbor (NN) to match the descriptors, (2) hypothesis generation, obtaining a transformation from the object model to the possibly detected object in the scene, and (3) verification, to determine if the obtained transformation is valid for the model and the hypothesis.

In the literature, grasping methods are classified into two categories (Bohg et al., 2014): analytical and empirical. In the analytical approaches, physical formulations are applied in order to synthesize grasp points (Bicchi and Kumar, 2000). Otherwise, in the case of empirical solutions, mathematical and physical models are used allowing the system to learn from simulations or a real robot (Kamon et al., 1996).

This work is focused on empirical approaches in order to automatically obtain grasp points by using 3D registered object models. A grasp point is normally parametrized by the point on which the center