In Search of 3D Poses: Advancing Feature Description, Matching and Estimation for Robotic Applications

One of the key challenges for realizing autonomous robots is the perception problem. The “eyes” of a robot consists of one or more sensors, for example color cameras, which enable the robot to “see” its environment. Over the recent decades, another type of sensor has become popular: the range sensor. A range sensor adds a third dimension of depth, which allows the robot to sense 3D structures. This makes manipulation tasks, such as reaching out for a cup, easier to solve, since the depth can be determined from these structures. However, for any of this to be possible, the robot has to be able detect the location of the cup within the sensed 3D environment, and this is a challenging task, especially in real, chaotic environments, which contains a high amount of disturbance factors, e.g. other objects, backgrounds, and foreground elements occluding the object.. In industry, robots are often used for simple pick and place tasks, with objects being fixated in the same position for the robot to be able to grasp them. If the robot would automatically detect and locate the objects, fixtures would not be required, and the robot system would be more flexible and “intelligent”, because it would allow for easier adaptation to new scenarios with new objects.

This PhD thesis deals with automatic detection and localization of objects in 3D data. The starting point is a 3D input from a range sensor, containing depth information for the visible area in the environment, also called a “scene”. This could be a kitchen environment containing a table with one or more cups on it. The task is to locate exactly the portion(s) of the scene containing the relevant items for the robot. Since scenes represented by a range sensor often contain hundreds of thousands of 3D points, the search for an object within this data is a complex task.

In this thesis, the problem is attacked by the use of local 3D feature representations. Local 3D features are used for describing small parts of an object/scene in terms of the variation of the surface. Distinct parts can then be matched, providing evidence of the presence of the object. An example could be a local feature describing a handle, which is a strong indicator of a cup being present. On the other hand, a curved surface patch is a weak indicator, since it can come from several different objects or containers. The matching problem consists of finding associations between these local structures in an intelligent manner. The thesis presents new methods for performing this matching, by which weak matches are removed. The final stage is the localization process, which is based on the associations made between local structures on the object and in the scene. The localization methods generally work by virtually placing the object in the scene according to one or more matches between local structures. For instance, the object-scene association for a handle can be used to create a virtual or simulated scene containing the object. Based on this hypothesis placement, the plausibility of the virtual structures are tested. In this last, crucial step, wrong localizations are rejected. As an example, a mismatch can cause impossible placements of the object, e.g. with parts of the cup going through the table. The thesis introduces two methods for efficiently generating plausible hypothesis placements, thus leading to faster and better localizations of objects.

The sum of our contributions will motivate further research on realizing more intelligent robotic agents for e.g. grasping and meaningful manipulations of objects.