Active strategies for object discovery

Phil Bradfield and Jan Fabian Schmid

Master Project in Computer Vision, Fachbereich Informatik, Universität Hamburg

Motivation

Human environments are full of significant objects - from pens, to coffee mugs, to computer monitors - which a mobile robot system might need to find, examine, manipulate or otherwise interact with. But how can the robot identify what is an object and what is not? And, when dropped into an unknown environment, how can it intelligently explore in order to increase its knowledge of the locations and shapes of the objects in its vicinity as efficiently as possible?
This project investigated object detection and the "next best view" (NBV) problem: given the robot's current knowledge about its environment, where should it move to next in order to maximally increase its knowledge?

Objectives

- To develop software for a mobile robot to autonomously explore an unknown environment and discover objects within it.
- To use the system to compare different approaches to the NBV problem.

System overview

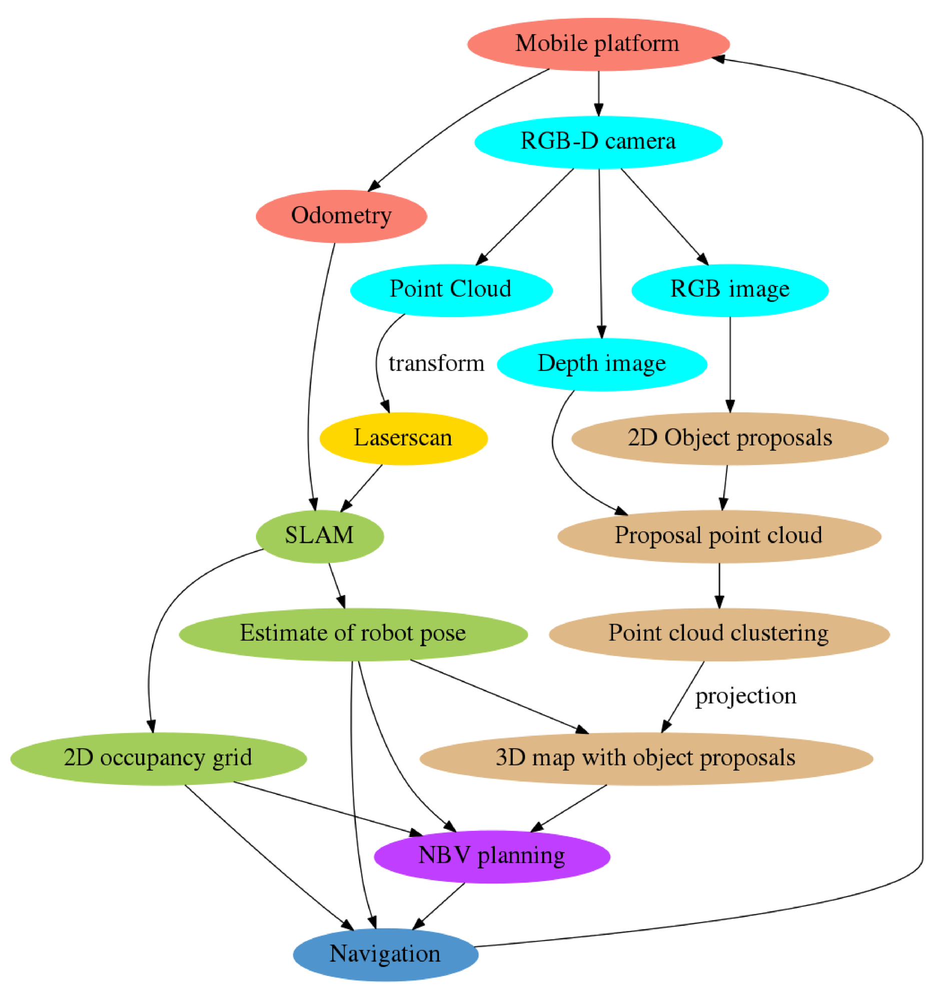

Figure: System overview

1. Pioneer 2 robot moves around the environment and provides odometry information (red).
2. Mounted Kinect camera provides RGB images, depth maps, and point clouds (light blue).
3. Point cloud is transformed to laserscan (yellow) and combined with odometry data to perform SLAM (green).
4. 2D object candidates are detected from the RGB image using the VOCUS2 saliency system[1] and projected into a 3D octomap[2] using the depth values. Overlapping candidates are merged into coherent single entities (brown).
5. Map of the environment, estimated robot pose and object candidates are used to calculate the next best view pose (purple).
6. The NBV to move to is sent to the navigation system (dark blue) which calculates a path and provides motor instructions to the Pioneer.
7. System framework and inter-module communication implemented using ROS

Experimental setup

Experiments were conducted in the Gazebo simulator using the YCB object set[3]. Five NBV calculation methods were implemented:

- `Random` choice
- Frontier-based exploration (`FBE`), which seeks to simply view all unseen space as quickly as possible
- Sampling-based exploration, which samples the known free space and calculates the information gain at each sampled point based on only known obstacles (`SBE (obs)`) or both obstacles and candidates (`SBE (obs+cands)`)
- Frontier+ (`FBE plus`), which uses `FBE` in the initial exploration phase, then switches to `SBE (obs)`

Sampling-based methods calculated information gain by raycasting from each candidate viewpose, and scoring the pose based on the types of voxels visible from it: unknown voxels scored 2 points, obstacles 10, and unknown voxels next to objects 20. Additionally, SBE (obs+cands) awarded extra bonuses for voxels containing candidates and unknown voxels next to candidates.

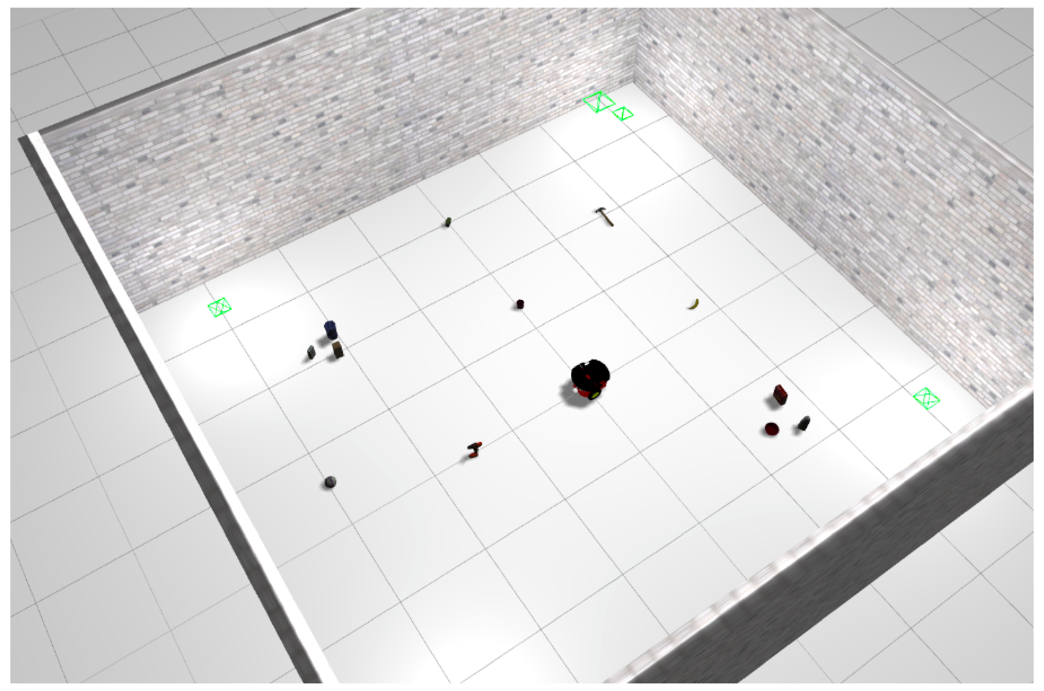

Figure: Experimental scenario

Results

- `FBE` was fastest to explore the environment
- `SBE (obs)` had best overall performance, with recall of 0.75 and precision of 0.25 after 25 views
- `FBE plus` implemented to combine the best of these two methods
- `SBE (obs+cands)` failed to improve on the performance of `SBE (obs)`; presumably because precision of detections too low for extra candidate information to improve performance

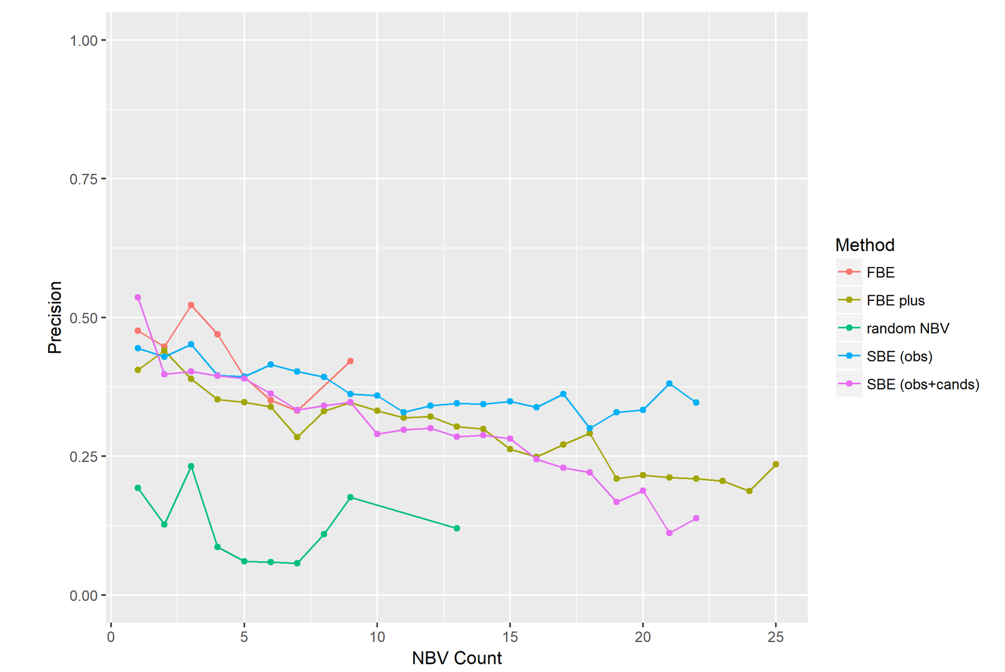

Figure: Development of object-level precision for different NBV methods

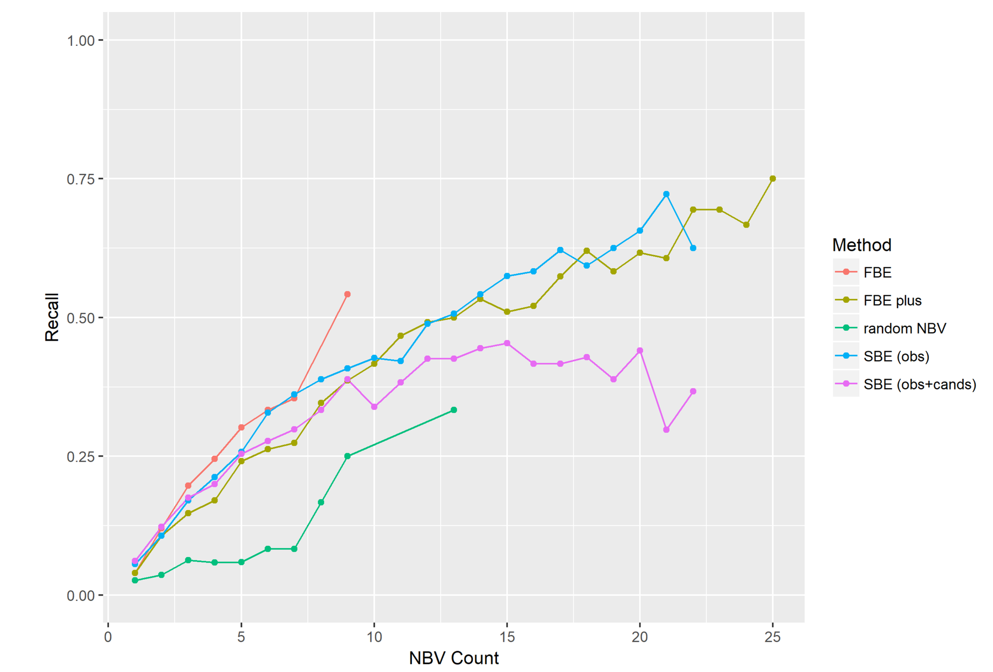

Figure: Development of object-level recall for different NBV methods

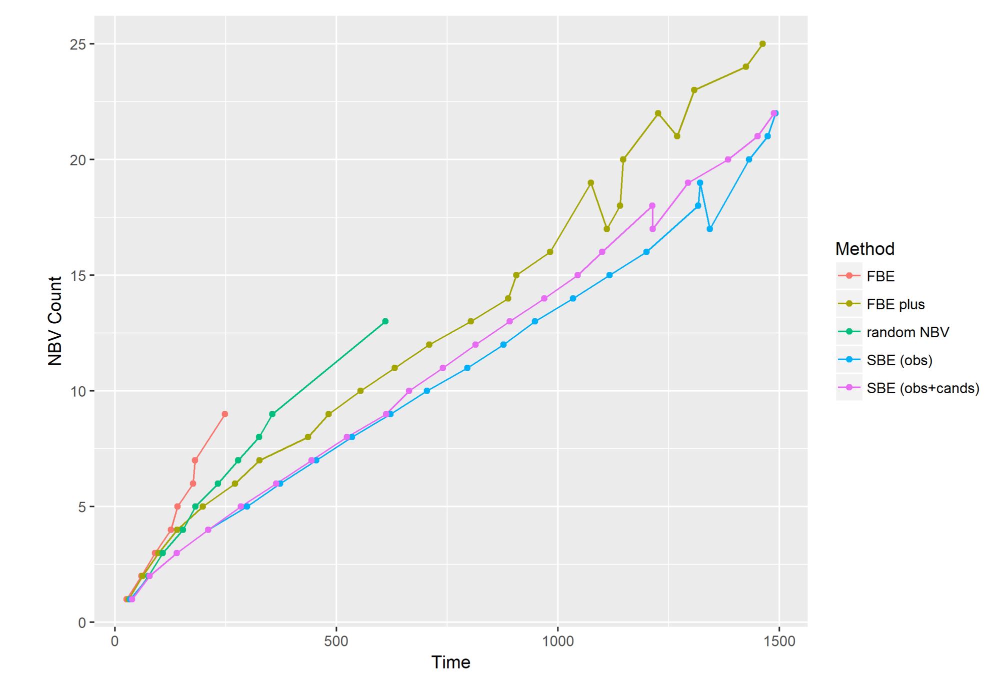

Figure: Time taken for different NBV methods to take views

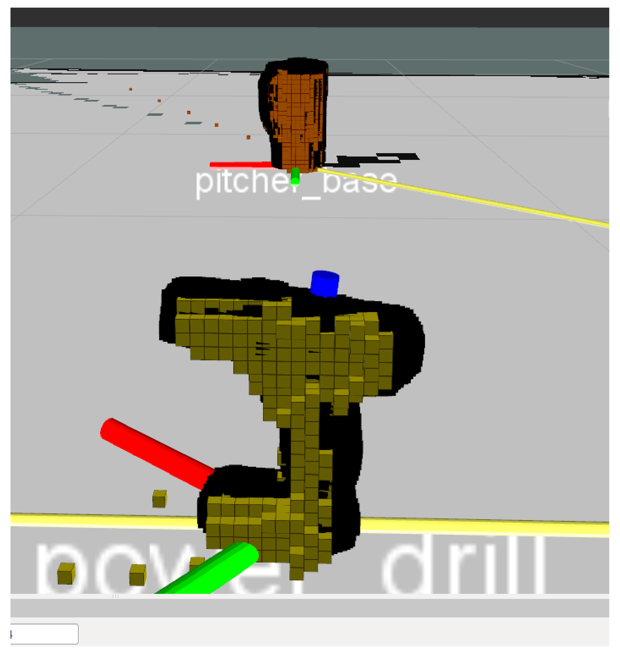

Figure: Example object candidates (coloured voxels) which match closely to the ground truth (black point cloud)

Possible extensions

- Improve SLAM system
- Include depth information in object candidate computation
- Use heuristics to estimate the shapes of unseen areas of objects

References

[1] Germán Martín García and Simone Frintrop.
A Computational Framework for Attentional 3D Object Detection.
In Proc. of the Annual Conf. of the Cognitive Science Society, 2013.

[2] Armin Hornung, Kai M. Wurm, Maren Bennewitz, Cyrill Stachniss, and Wolfram Burgard.
OctoMap: an efficient probabilistic 3D mapping framework based on octrees.
Autonomous Robots, 34(3):189–206, Apr 2013.

[3] B. Calli, A. Walsman, A. Singh, S. Srinivasa, P. Abbeel, and A. M. Dollar.
Benchmarking in Manipulation Research: Using the Yale-CMU-Berkeley Object and Model Set.
IEEE Robotics Automation Magazine, 22(3):36–52, Sept 2015.

Acknowledgements

The system was implemented as part of the module Masterprojekt Computer Vision, run by the Computer Vision group in Fachbereich Informatik at Universität Hamburg.

Contact Information

- Email:{Philip.Bradfield,Jan.Fabian.Schmid}@informatik.uni-hamburg.de
- The CV Group: https://www.inf.uni-hamburg.de/en/inst/ab/cv.html