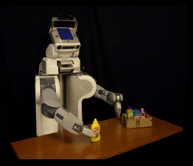
PERCH: Perception via Search for Multi-Object Recognition and Localization

Carnegie Mellon University
THE ROBOTICS INSTITUTE

Venkatraman Narayanan and Maxim Likhachev

Problem Statement

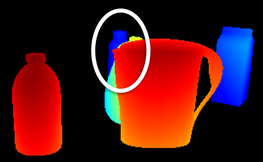



task

identify type and 3 DoF pose of every object in the scene (point cloud/depth image)

given

6 DoF camera pose,
3D models of objects in the scene



Feature and template-based methods are brittle (e.g., occlusion)

Learning methods need training data to capture the combinatorics of inter-object interactions

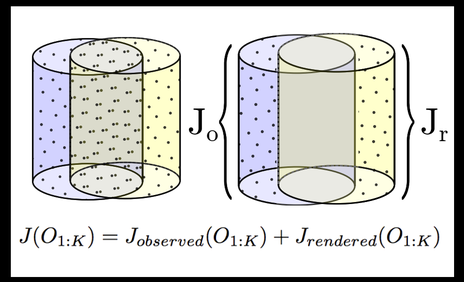
Contributions

- Deliberative perception: search for best "hypothesis " over space of rendered scenes
- Tree search decomposition for optimization: "Monotone Scene Generation Tree"
- Theoretical guarantees on bounded suboptimality of solution

Optimization Formulation

minimize cost

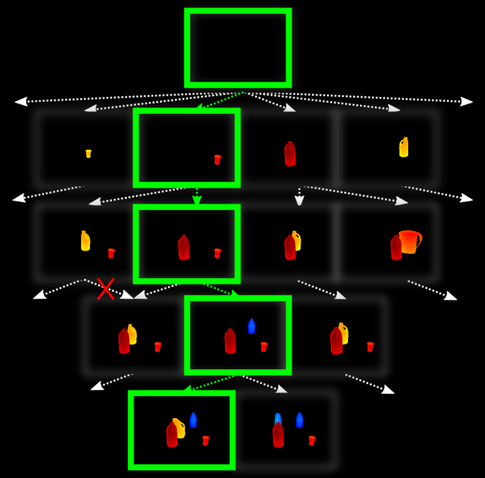
#unexplained points in observed cloud
+
#unexplained points in rendered cloud



Render all possible scenes, select the one that "best matches" the input depth image

- Brute-force search over joint state space is intractable: 4 objects, 10 (x,y) positions, 10 orientations: 100^4 states ~ 12 days on a GPU
- **Key idea**: cost function can be decomposed over objects under a monotonicity constraint: assign object poses sequentially; ensure penalty accrued for an assigned object never decreases later
- Constraint results in "non-occluding" order, problem reduces to tree search on this **Monotone Scene Generation Tree**

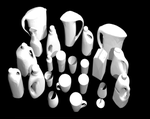
Monotone Scene Generation Tree & Heuristic Search

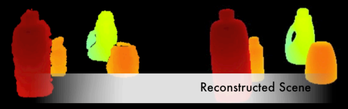


Find shortest path from root node to *any* leaf

- Tree search is still hard. For branching factor of 8000 and tree depth of 4, we have ~ 4×10^{15} nodes in the tree
- We use Focal Multi-Heuristic A* (MHA*)[1] for the search, and parallelize child node generation
  - Heuristic 1: prefer expanding nodes lower in the tree
  - Heuristic 2: prefer expanding nodes where assigned objects have maximum overlap with input point cloud
- Focal MHA* guarantees that solution is within desired quality bound, despite using arbitrary heuristics

Evaluation

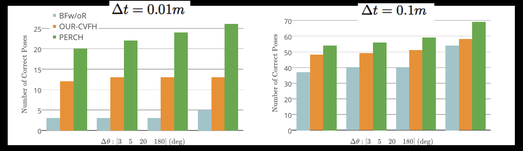




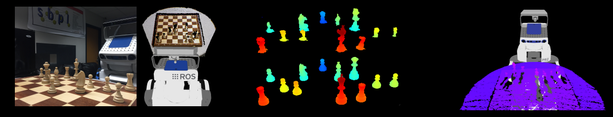
dataset

- Household objects occlusion dataset[2]
- 36 objects models, 82 instances in 23 scenes
- Discretization: 4 cm, 22.5 deg; ICP at every stage to compensate for discretization artifacts
- Parallel child node generation (AWS m4.10x, 2x40 cores)
- Mean search time: 6.5 mins

baselines
OUR-CVFH[3]
Brute-Force ICP



scaling experiment



Upcoming at Robotics: Science and Systems (RSS) '16
Discriminatively-guided Deliberative Perception (D2P)

code: github.com/venkatrn/perception

[1] Improved Muti-Heuristic A*, Narayanan et al., SoCS '15
[2] Point Cloud Library, Aldoma et al., IEEE RAM '12
[3] CAD Model Recognition and 6 DOF Pose Estimation using 3D Cues, Aldoma et al., ICCV Workshops, '11

This research was sponsored by ARL, under the Robotics CTA program grant W911NF-10-2-0016

venkatraman@cs.cmu.edu