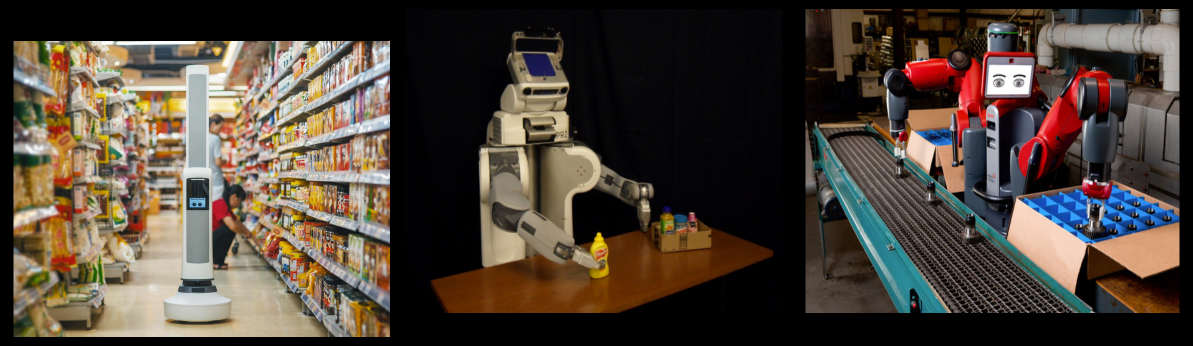
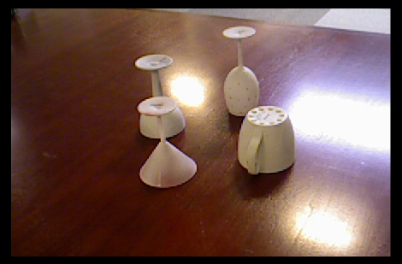
Discriminatively-guided Deliberative Perception (D2P) for Pose Estimation of Multiple 3D Object Instances

Venkatraman Narayanan and Maxim Likhachev

Carnegie Mellon University

Problem Statement



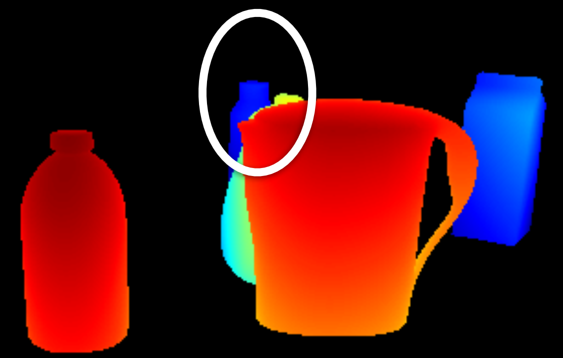


task:

identify type and 3 DoF pose of <u>all</u> objects in the scene (point cloud/depth image)

given:

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



- Feature and learning based methods are brittle (e.g., occlusion)
- Learning methods need training data to capture the combinatorics of inter-object interactions
- Generative methods are robust but slow

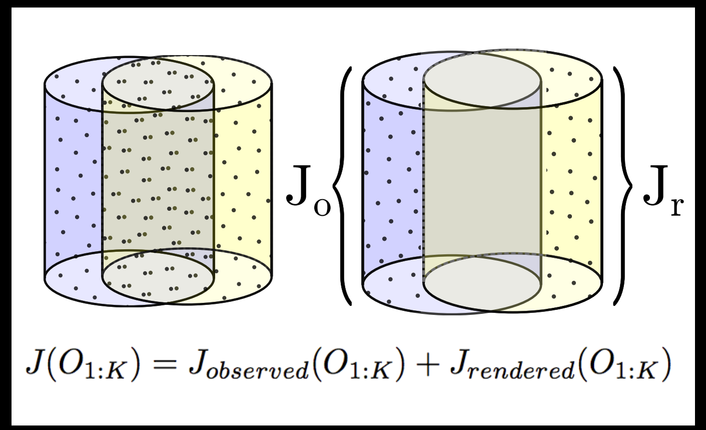
Contributions

- Framework to guide deliberative global search with *any* and *multiple* discriminative learners
- Theoretical guarantees on solution quality
- Notion of completeness for multi-object recognition and pose estimation
- Lazy Multi-Heuristic A*

Perception via Search (PERCH)[1]

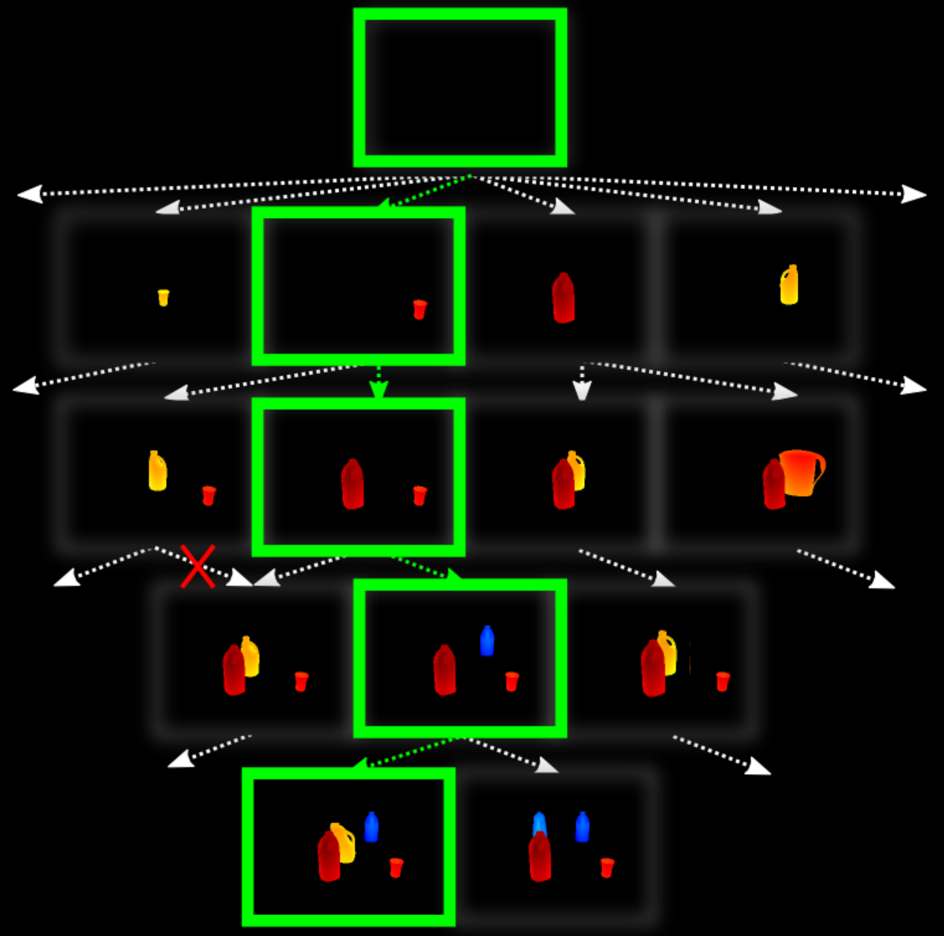
minimize over joint object poses

unexplained points in observed cloud +
unexplained points in rendered cloud



Render all possible scenes, select the one that "best matches" the input point cloud

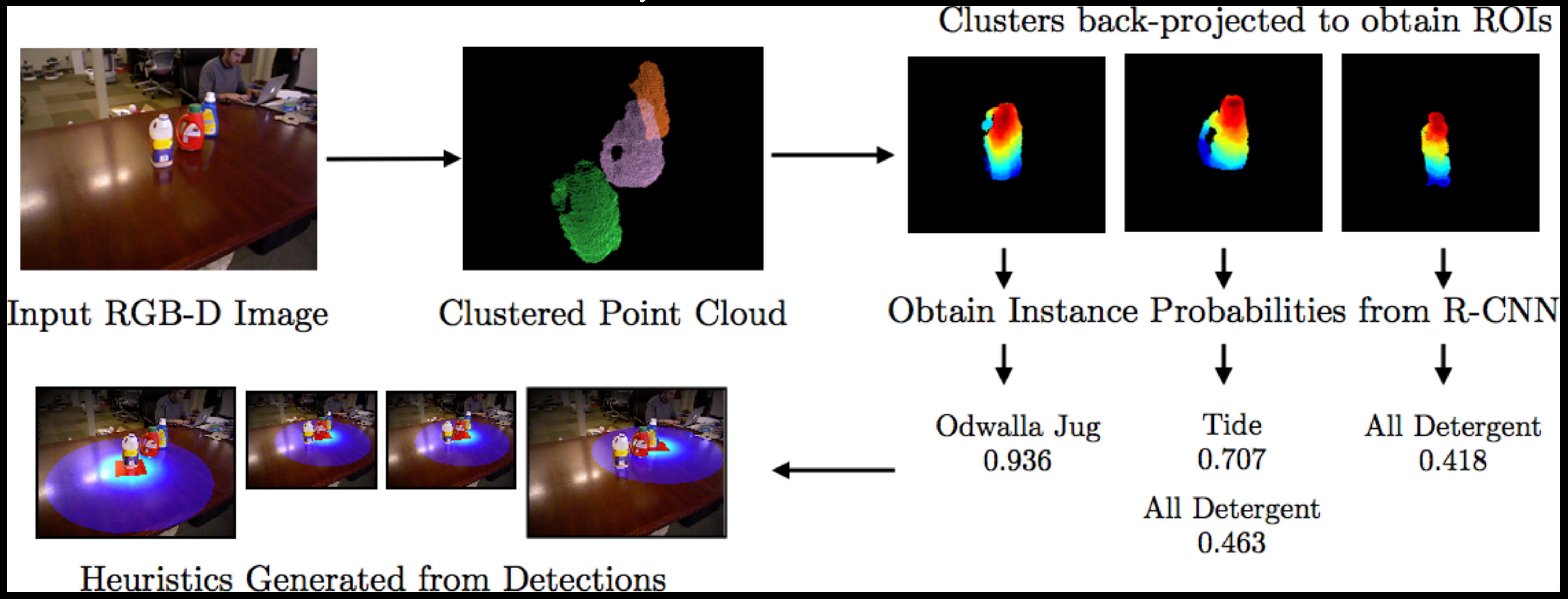
Brute force run time exponential in #objects (~12 days for 4 objects with 10 (x,y) positions, 10 orientations — 100^4 states)



Joint optimization can be cast as tree search, with monotone constraint! — still uninformed search

Key Idea: Guide deliberative search with discriminative learners

Generating Multiple Discriminative Heuristics



Discriminative heuristics can be inadmissible and unrelated to optimization objective
FOCAL Multi-Heuristic A* [2] was developed to handle precisely such heuristics

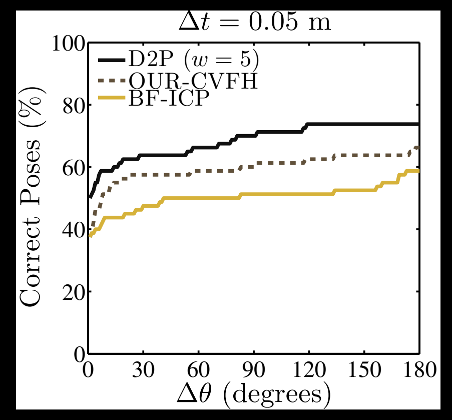
Completeness

Existing methods based on global hypothesis verification are "incomplete": can fail to produce a feasible solution (set of object poses) even if one exists

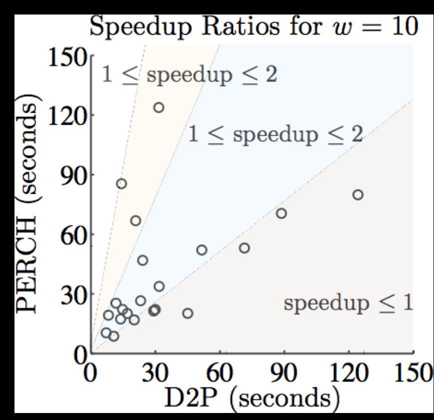
Evaluation

dataset

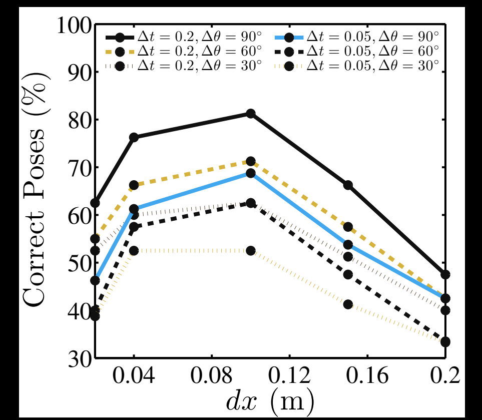
- Household objects occlusion dataset[3]
- 36 objects models, 82 instances in 23 scenes
- Discretization: 4 cm, 22.5 deg; ICP at every stage to compensate for discretization artifacts
- Parallel child node generation (AWS m4.10x, 40 virtual cores); lazy edge evaluation
- R-CNN network (ZF architecture) trained on synthetic depth images of individual objects



outperforms OUR-CVFH[4], BF-ICP



speedup over PERCH



discretization vs ICP tradeoff

code: github.com/venkatrn/perception

References

[1] Perception via Search Narayanan and Likhachev, ICRA '16
[2] Improved Muti-Heuristic A*, Narayanan et al., SoCS '15
[3] Point Cloud Library, Aldoma et al., IEEE RAM '12
[4] CAD Model Recognition and 6 DOF Pose Estimation using 3D Cues, Aldoma et al., ICCV Workshops, '11