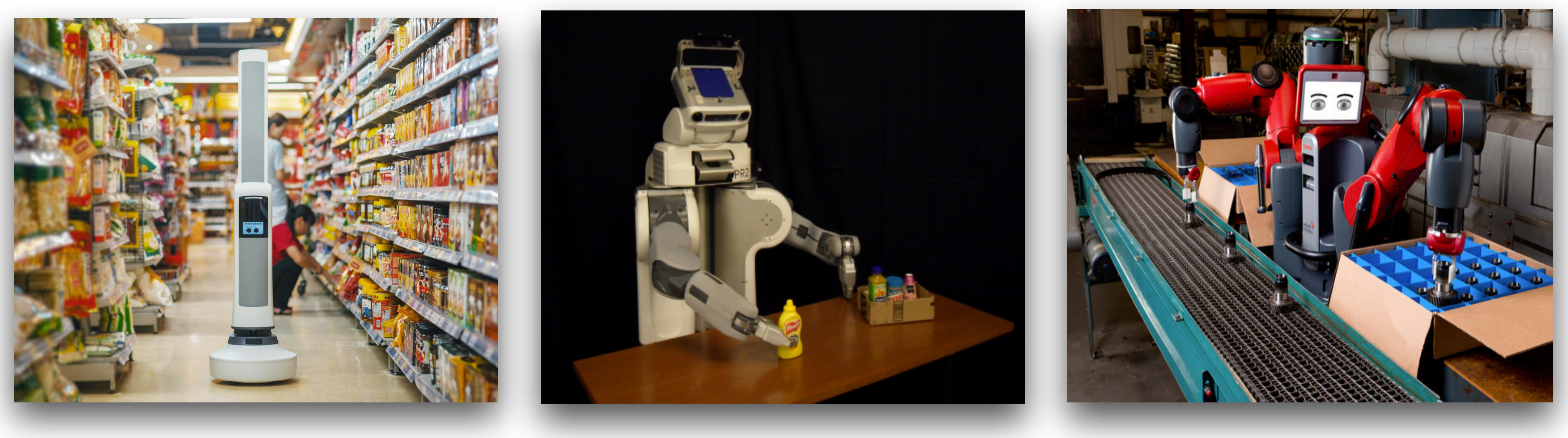
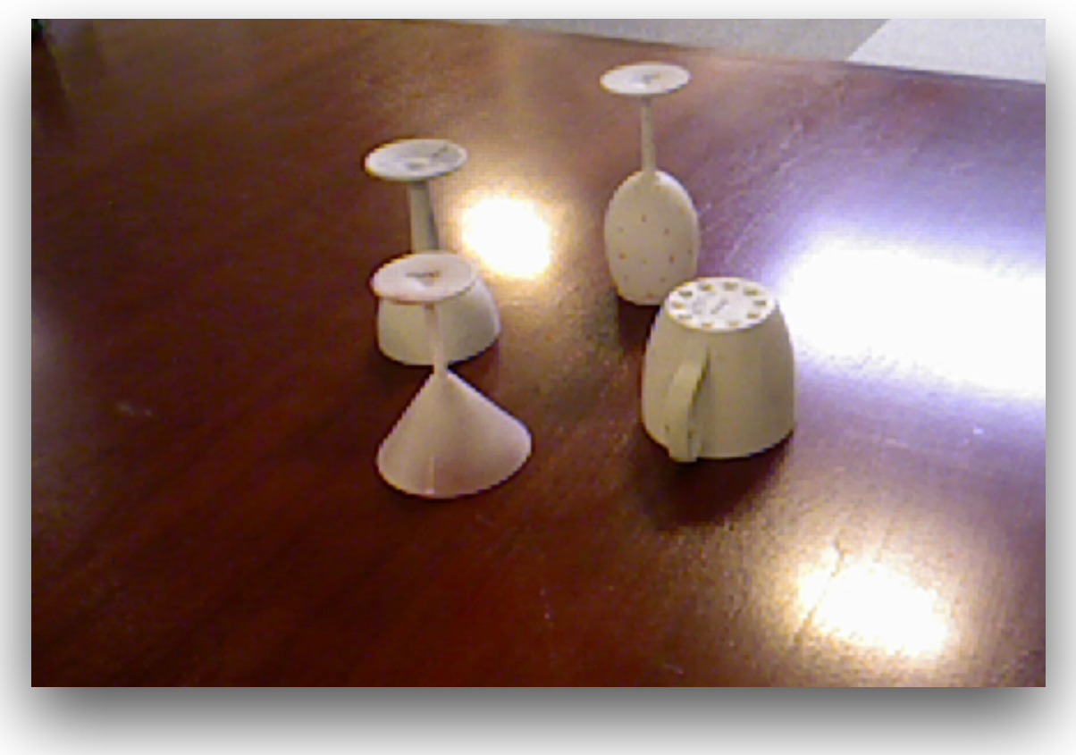
PERCH: **Per**ception via **Search** for Multi-Object Recognition and Localization

Venkatraman Narayanan
Maxim Likhachev
Carnegie Mellon University
The Robotics Institute

Problem Statement





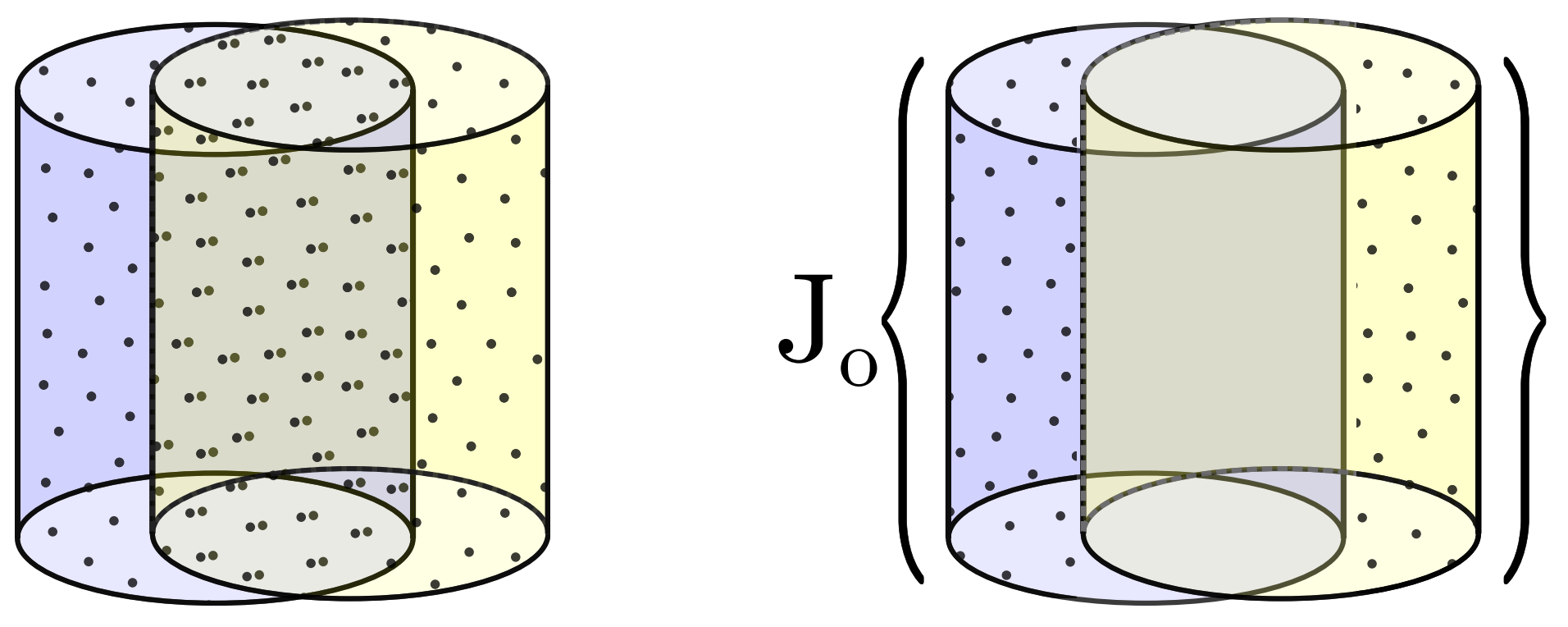
task
identify type and pose of every object in the scene (point cloud/depth image)

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

- Feature and template-based methods are brittle (e.g., occlusion)
- Learning methods need training data to capture the combinatorics of inter-object interactions
- We propose a deliberative approach that searches for the "best explanation"

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

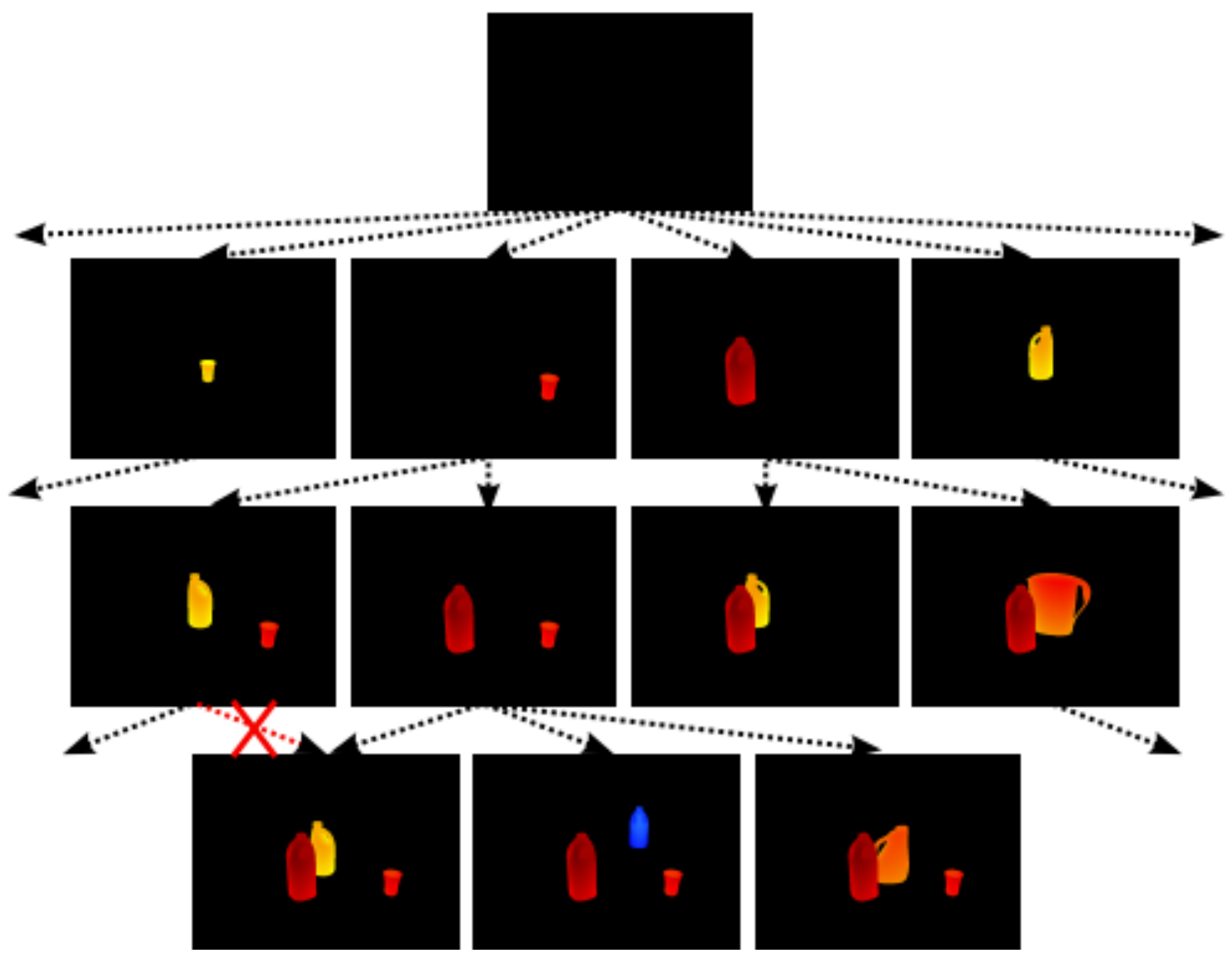
cost
#unexplained points in observed cloud
+
#unexplained points in rendered cloud



$$J(O_{1:K}) = J_{observed}(O_{1:K}) + J_{rendered}(O_{1:K})$$

Technical Details

- Brute-force search over joint state space is intractable: 4 objects, 100 (x,y) positions, 20 orientations: 2000^4 states
- **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**



cost for adding object at level *i* (edge cost)

$$\Delta J_r^i = \sum_{p \in \Delta R_i} \mathbb{1}_{[p \text{ is unexplained by } I]}$$

$$\Delta J_o^i = \sum_{p \in \{I \cap V(O_i)\}} \mathbb{1}_{[p \text{ is unexplained by } \Delta R_i]}$$

find shortest path from root node to *any* leaf node

- Tree search is still hard. For branching factor of 8000 and tree depth of 4, we have ~4 x10^{15} nodes in the tree
- We use Focal Multi-Heuristic A* (MHA*)[7] 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

Experiments

dataset

- Household objects occlusion dataset[2]
- 36 objects models, 82 instances in 23 scenes
- Objects vary only in (x, y, yaw)

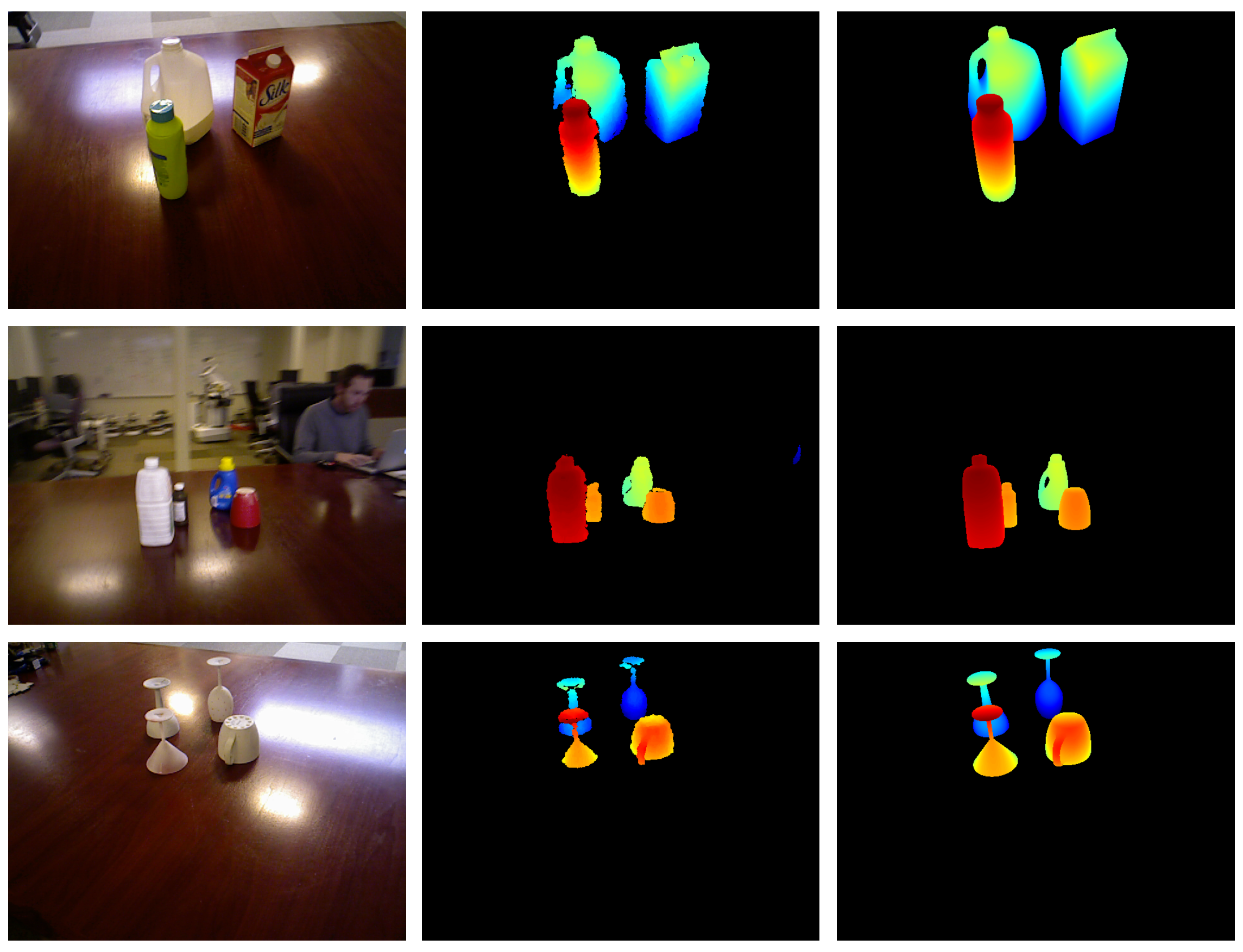
search configuration

- Discretization: 4 cm, 22.5 deg
- ICP at every stage to compensate for discretization artifacts
- Parallel child node generation (AWS m4.10x, 2x40 virtual cores)
- Mean search time: 6.5 mins

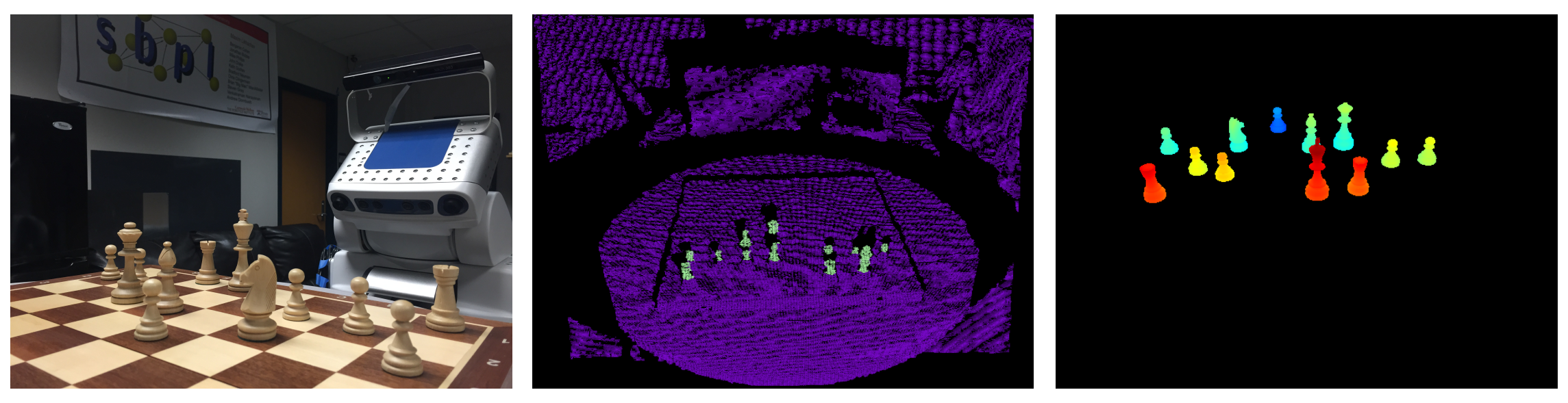
baselines

- OUR-CVFH[3]: Global viewpoint-based feature, robust to occlusions, trained with 642 views of every object
- Brute force ICP without rendering: slide every model over scene, take best fit over all possible orderings

RGB-D input **result**



scaling up

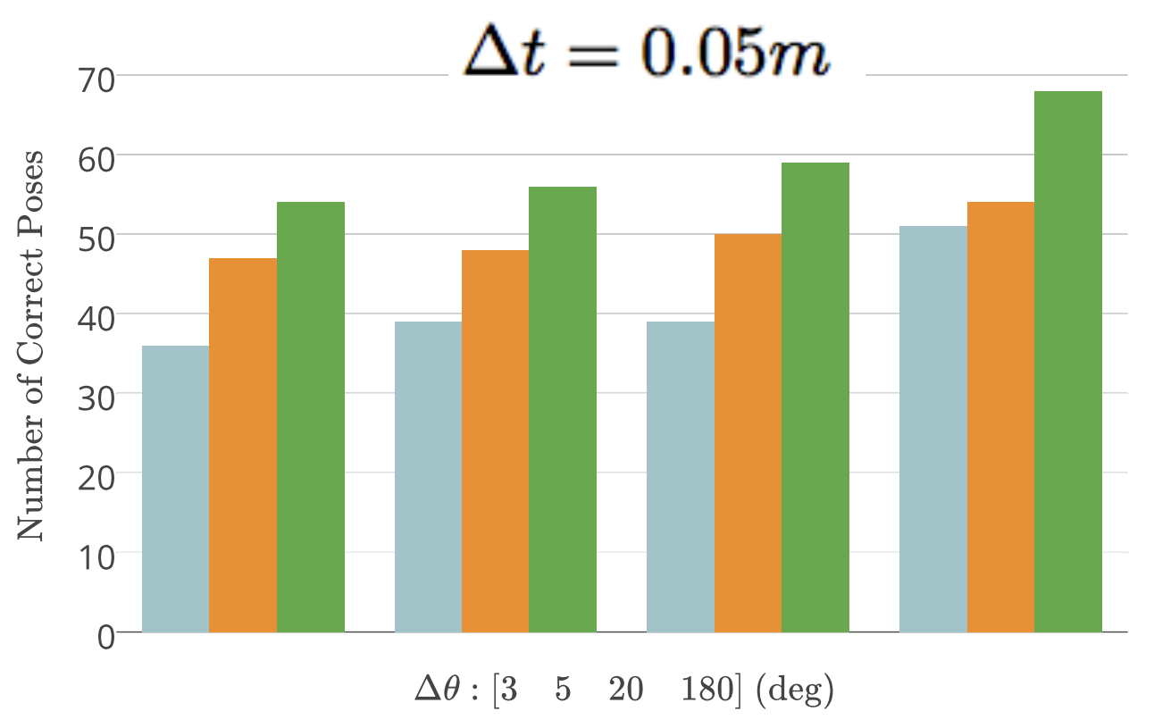


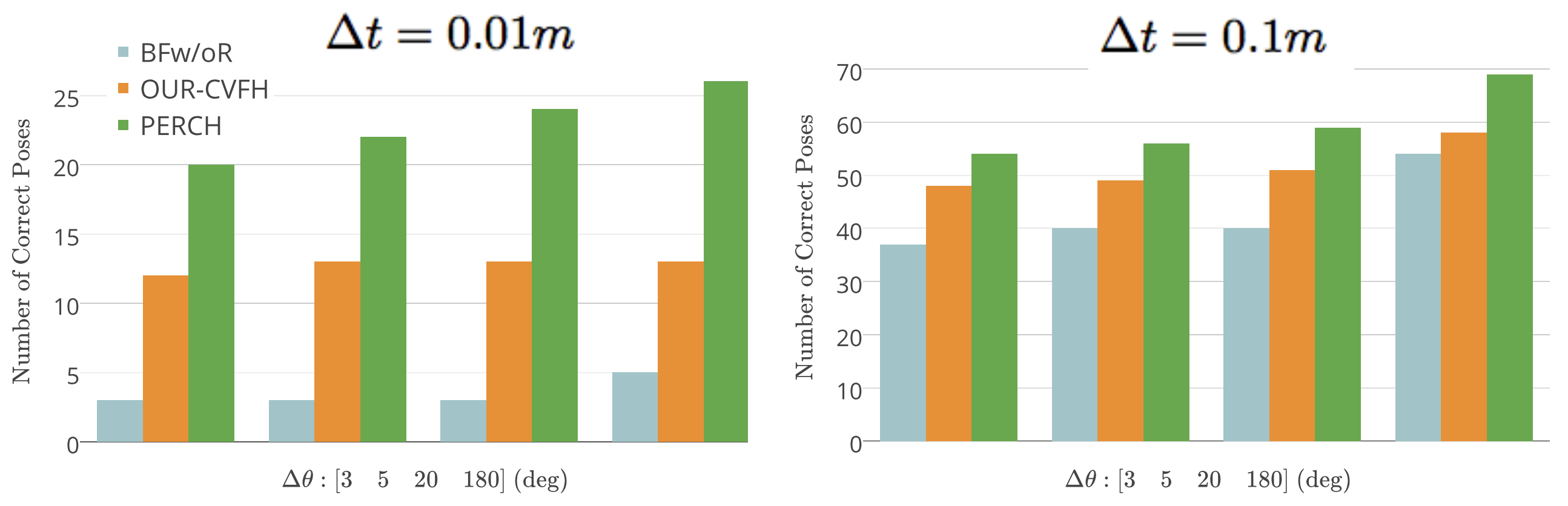
12 chess pieces, 6 unique models

pose correct if

$$\|(x,y) - (x_{\text{true}}, y_{\text{true}})\|_2 < \Delta t$$

$$\textsc{ShortestAngularDifference}(\theta, \theta_{\text{true}}) < \Delta\theta$$





[1] 3D ShapeNets, Wu et al., '15
[2] Point Cloud Library, Aldoma et al., '12
[3] OUR-CVFH, Aldoma et al., '12
[4] LINEMOD, Hinterstoisser et al., '12
[5] FPFH, Rusu et al., '06
[6] 3-D Shape Context, Frome et al., '04
[7] Improved MHA*, Narayanan et al., '15

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

venkatraman@cmu.edu