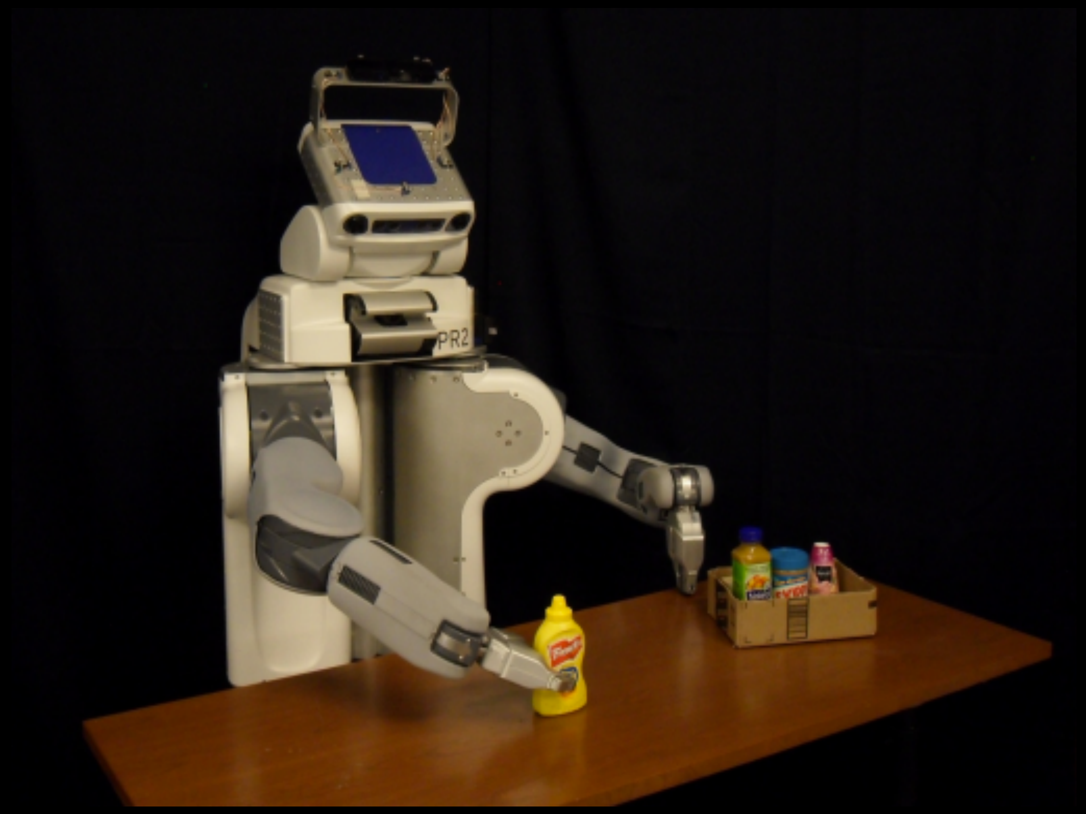
PERCH: Perception via Search for Multi-Object Recognition and Localization

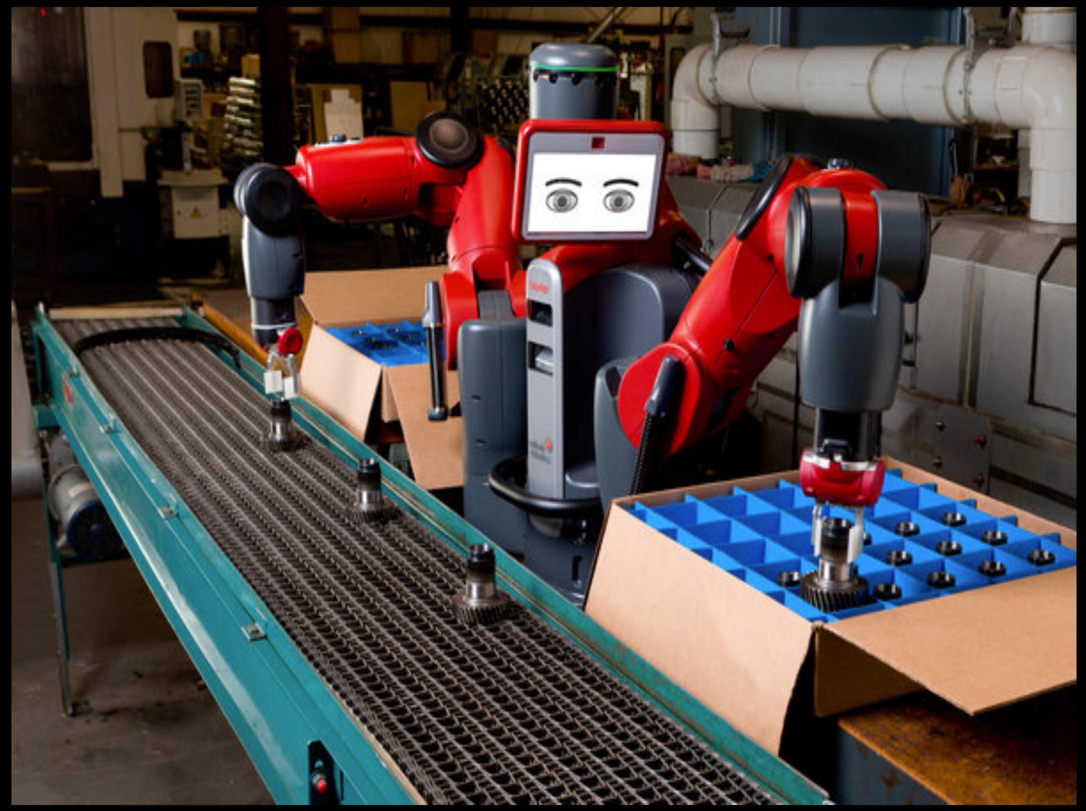
Venkatraman Narayanan and Maxim Likhachev

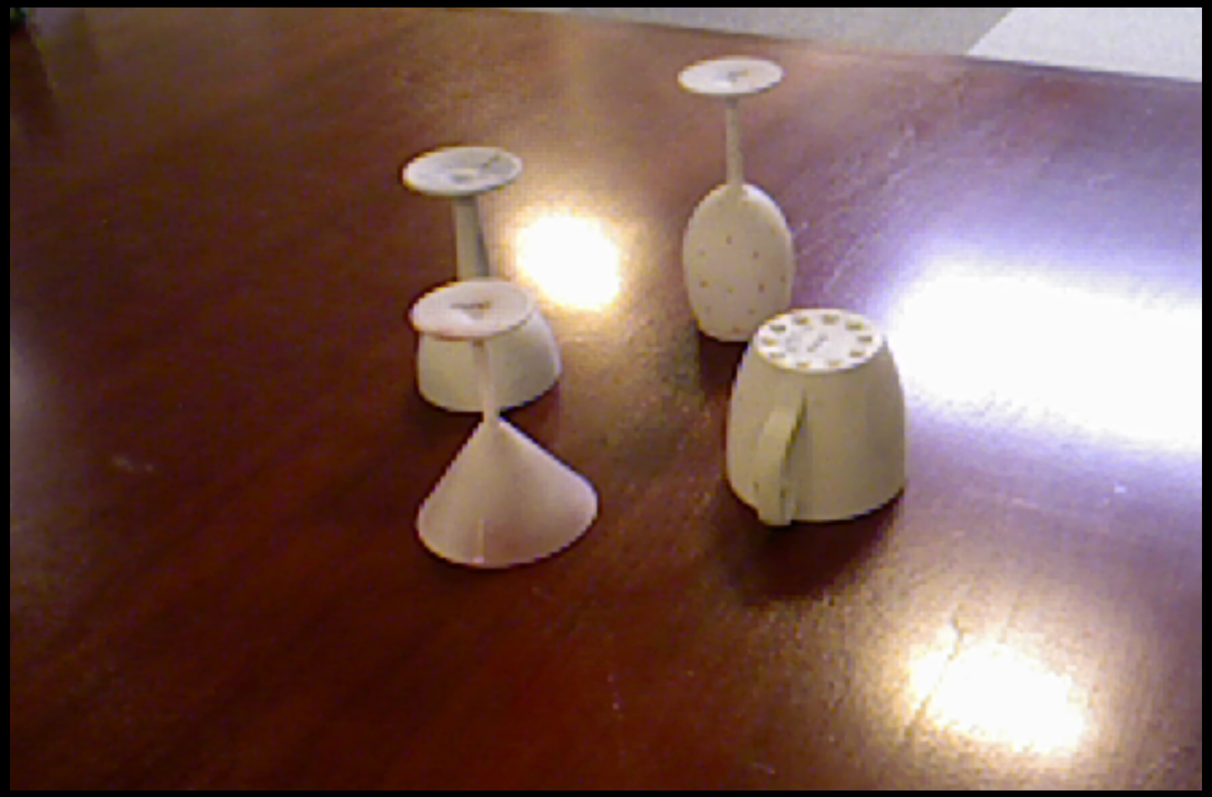
THE ROBOTICS INSTITUTE
Carnegie Mellon University

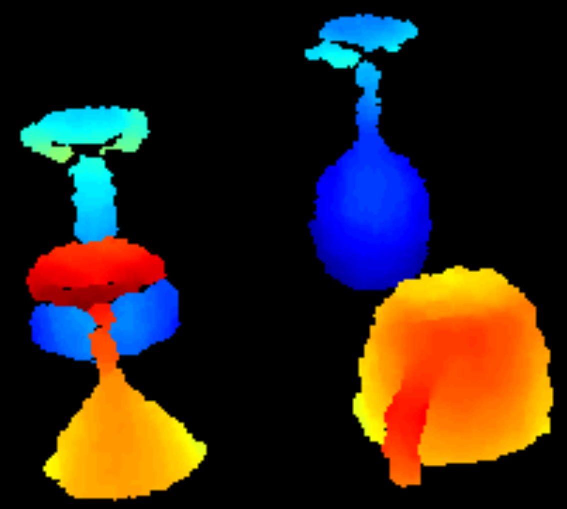
Image Courtesy: Simbe Robotics


Image Courtesy: Team HARP, CMU



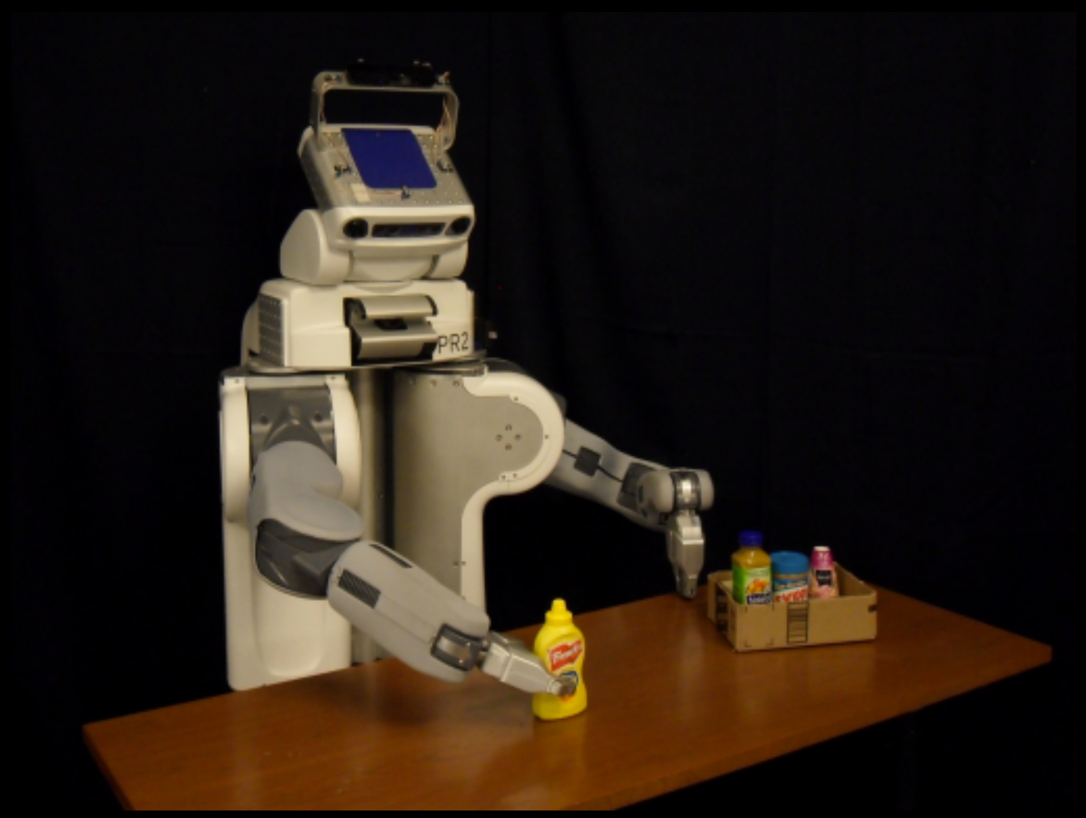


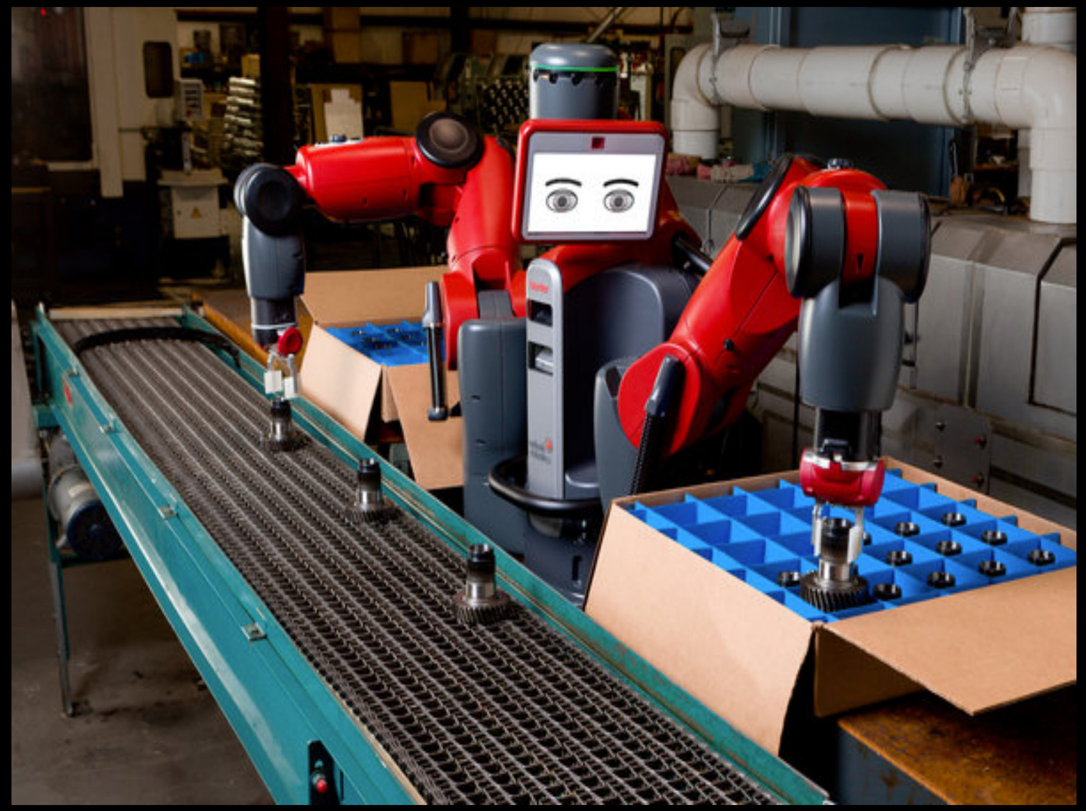


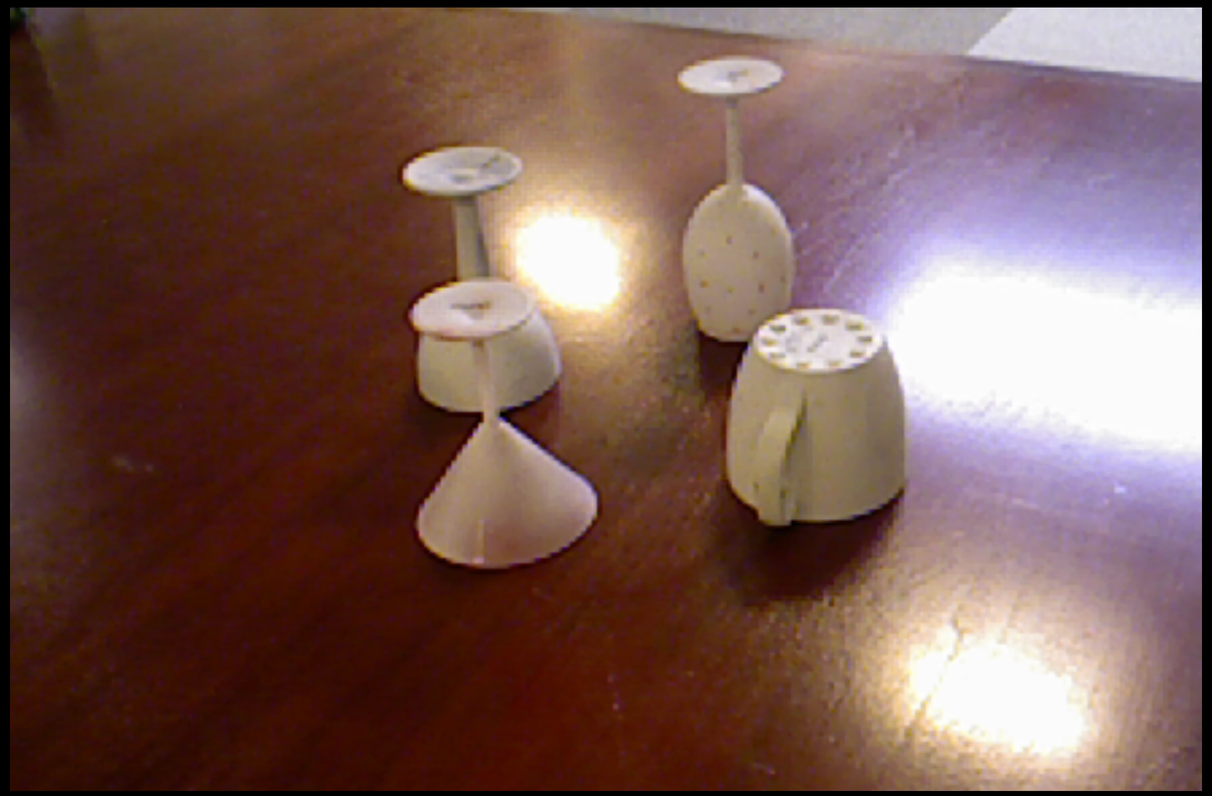


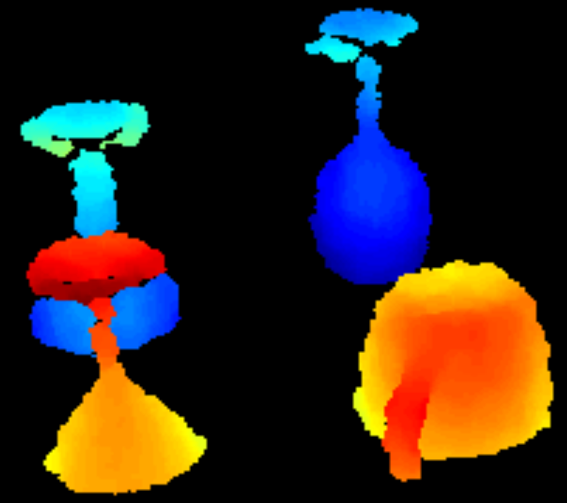
task: multi-object localization

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









task: multi-object localization

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

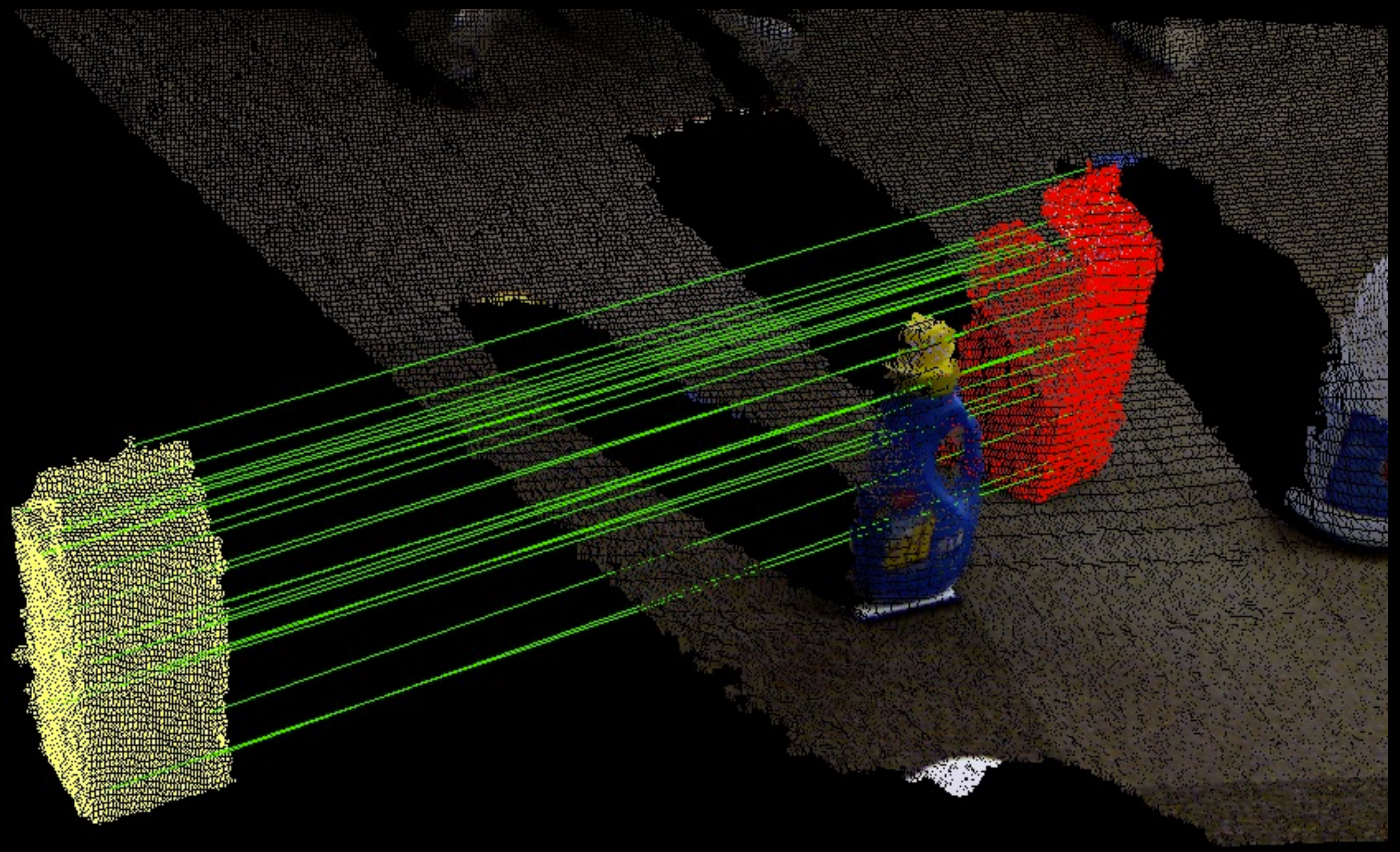
given

3D models of objects

6 DoF camera pose

Related Work

Local & Global Descriptor Matching



Spin Images, Johnson and Hebert, ’99

Radius-based Surface Descriptor, Marton et al., ’11

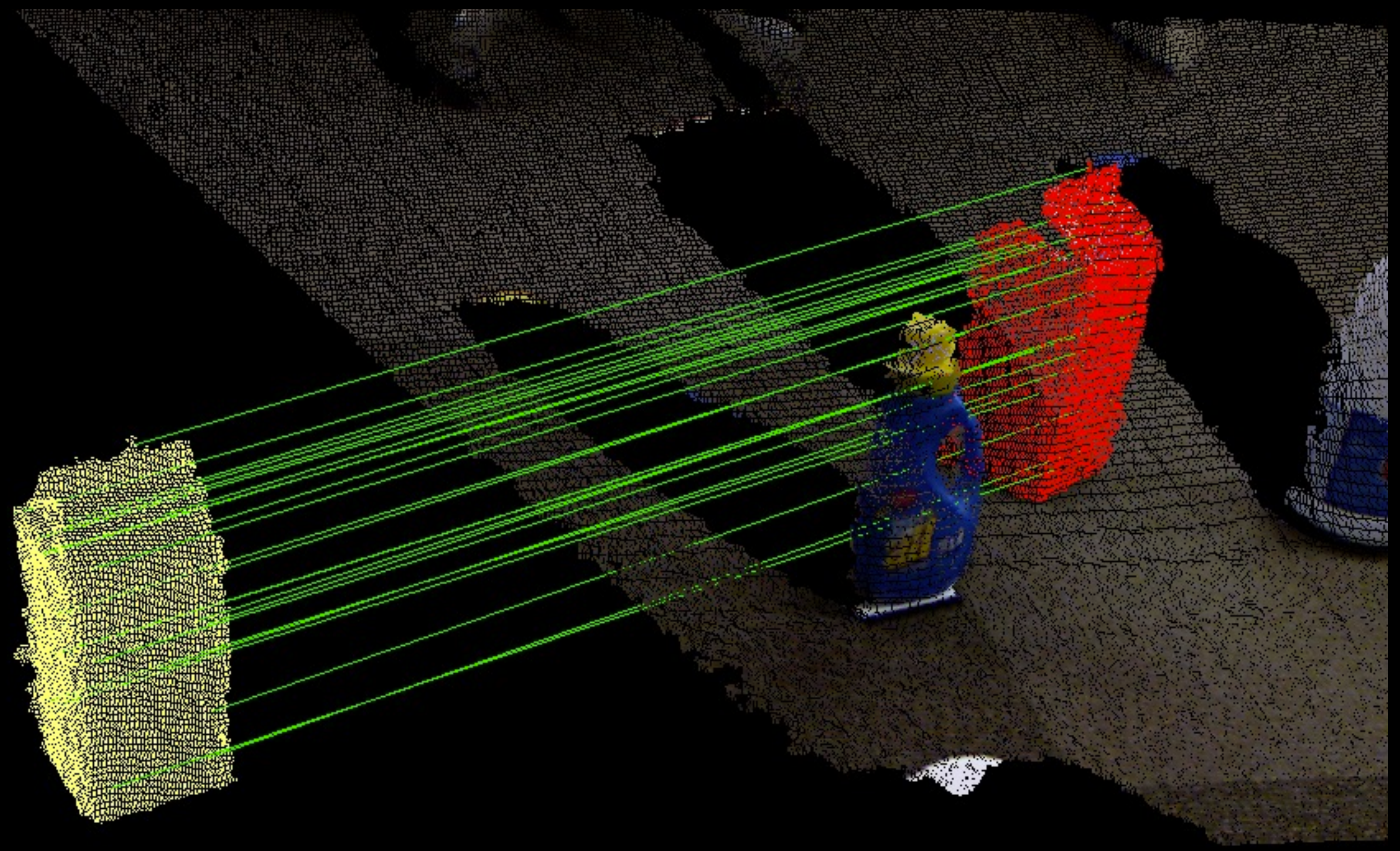
Viewpoint Feature Histogram, Rusu et al., ‘10

Clustered Viewpoint Feature Histogram, Aldoma et al. ‘12

Brittle, Sensitive to occlusions

Related Work

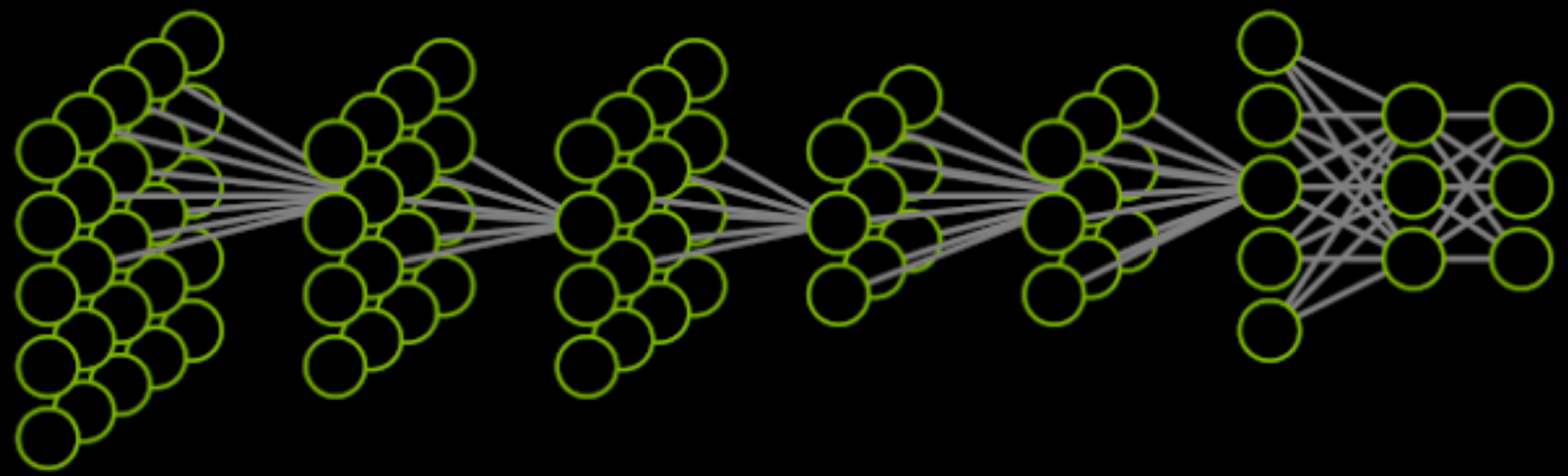
Local & Global Descriptor Matching



Spin Images, Johnson and Hebert, ’99
Radius-based Surface Descriptor, Marton et al., ’11
Viewpoint Feature Histogram, Rusu et al., ‘10
Clustered Viewpoint Feature Histogram, Aldoma et al. ‘12

Brittle, Sensitive to occlusions

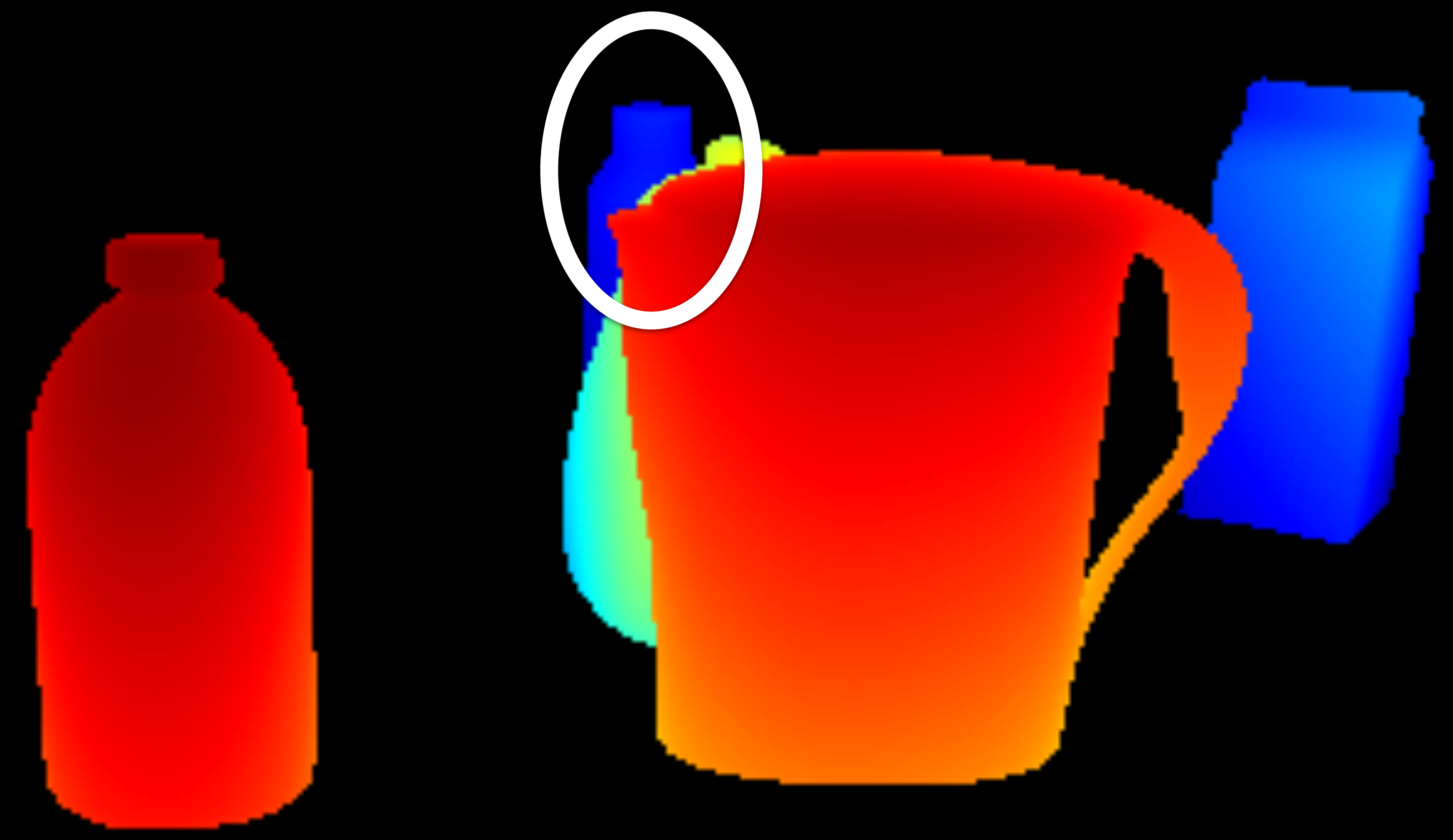
Learning-based Methods



Convolution-Recursive Network, Socher ’12
Transfer Learning, Alexandre, ’13
3D ShapeNets, Wu et al., ’15
VoxNet, Maturana and Scherer, ’15

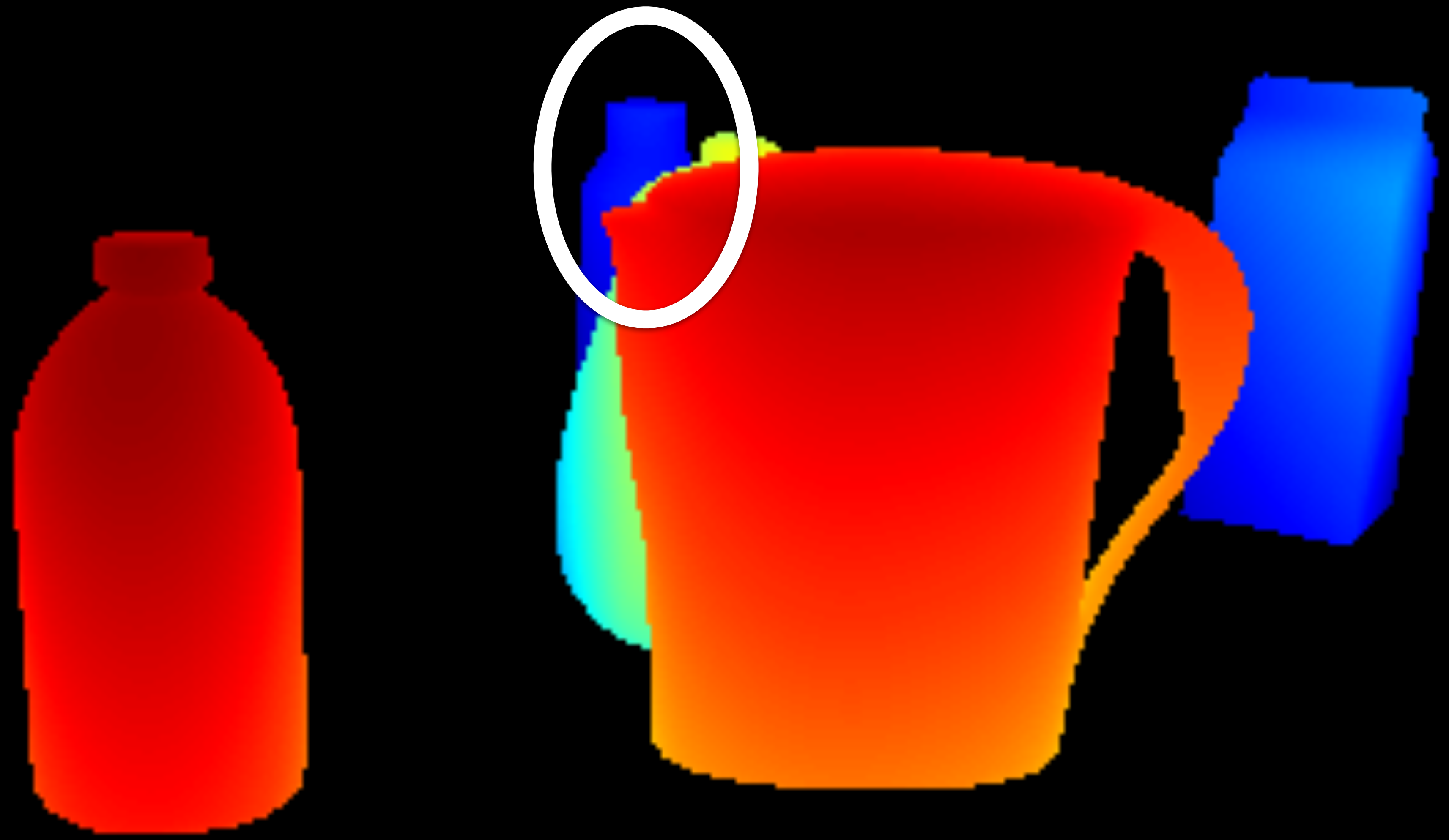
Combinatorial training data

Need for *Deliberative* Perception



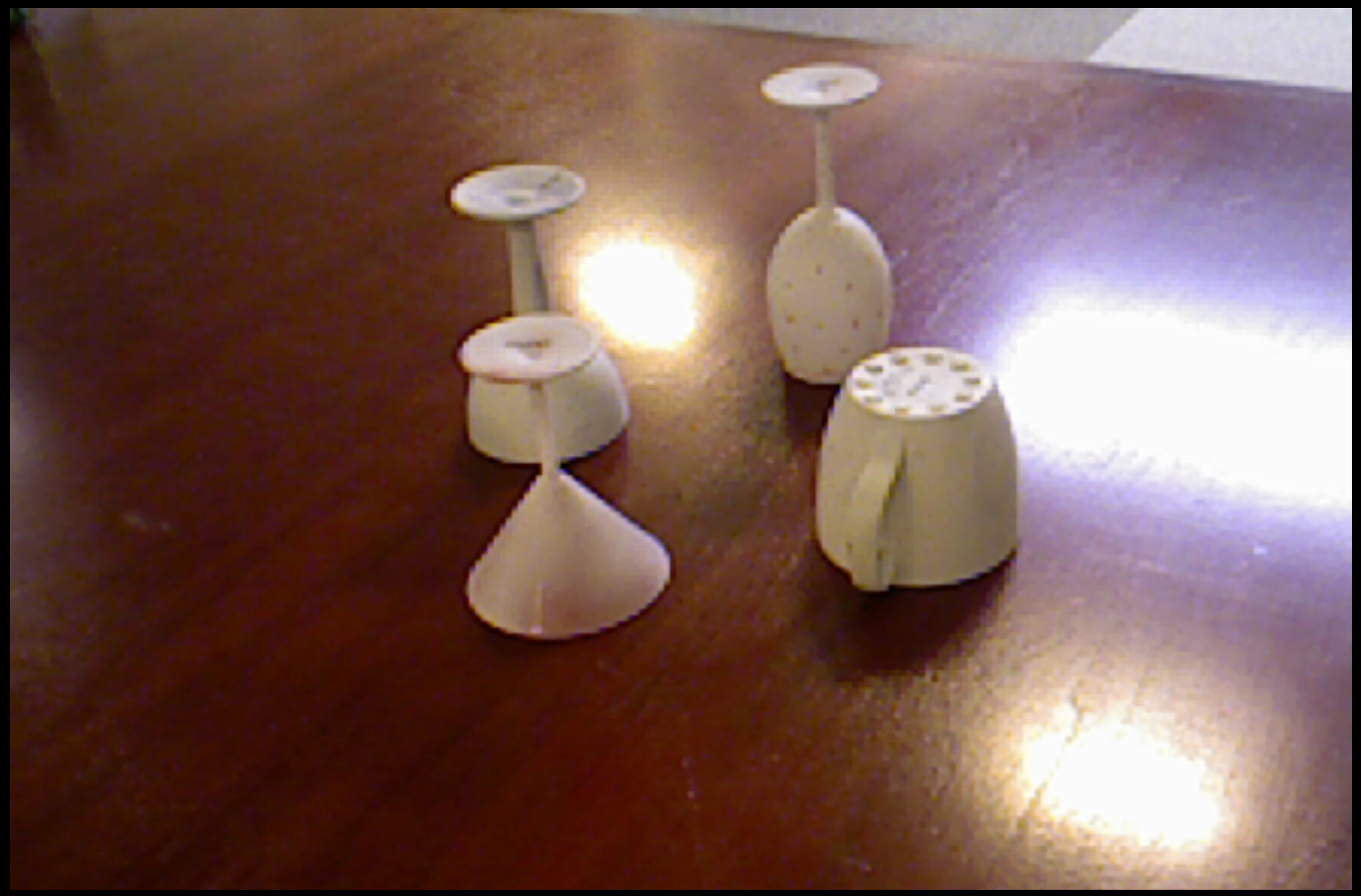
Need for *Deliberative* Perception

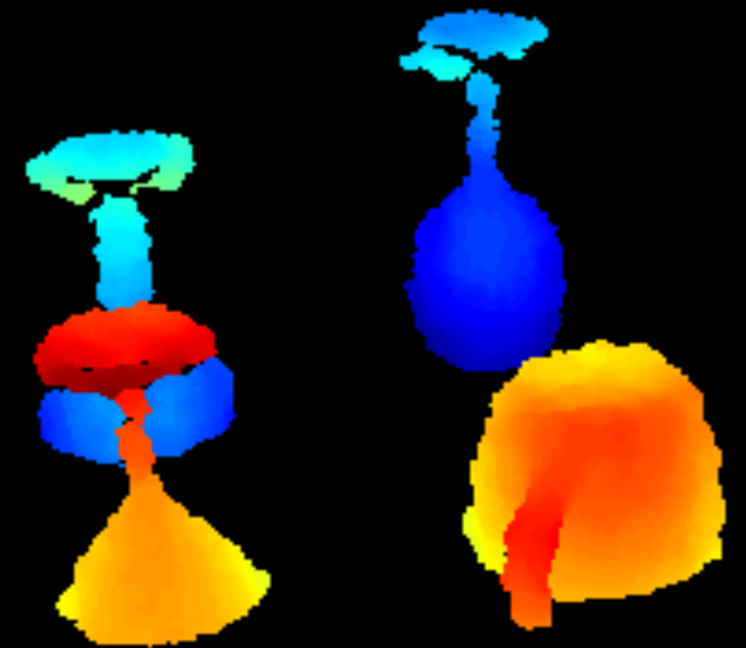
Explicit reasoning about inter-object occlusions



Contribution: Deliberative Perception

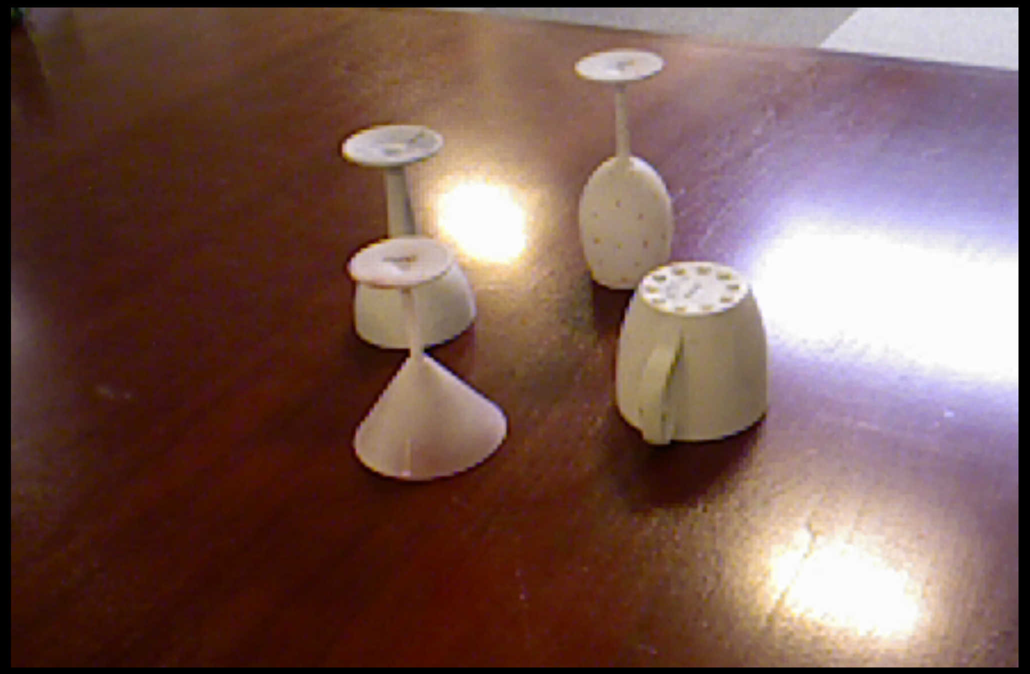
PERCH: Perception via Search

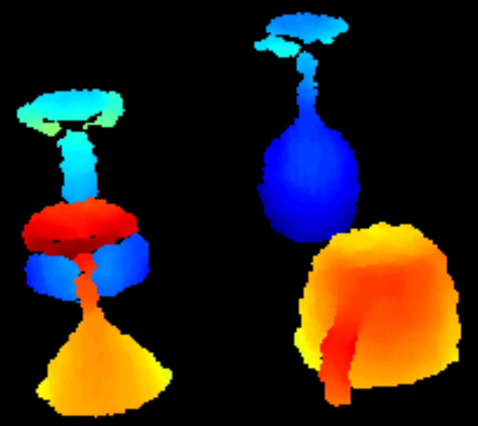




Optimize over space of all possible renderings
taking inter-object occlusions in account

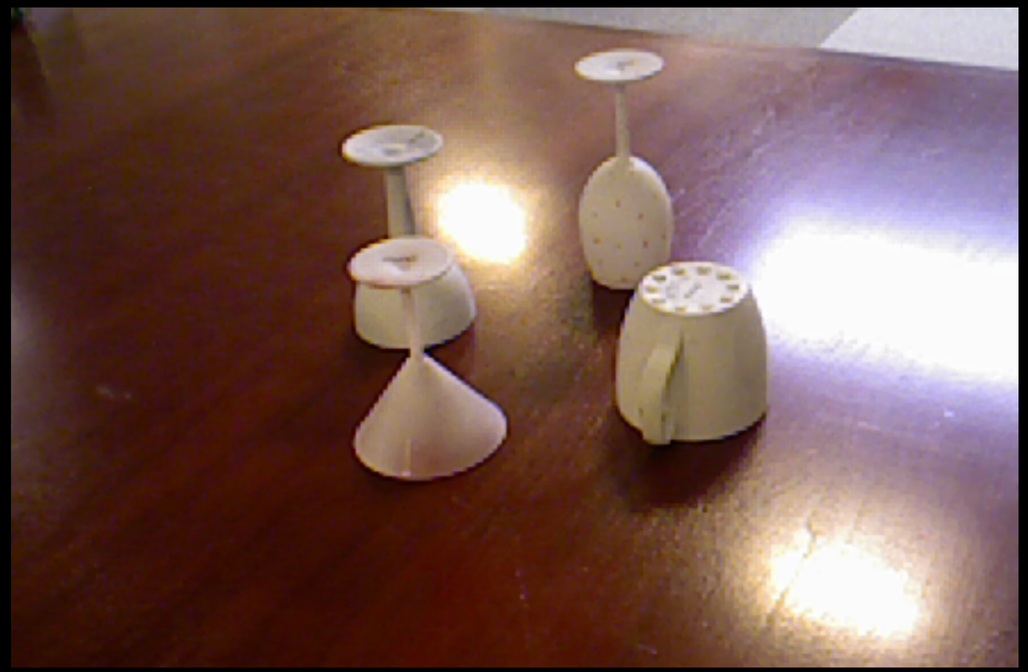
PERCH: Perception via Search

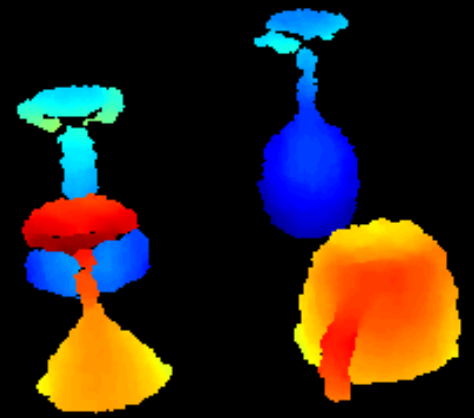




Optimize over space of all possible renderings

PERCH: Perception via Search





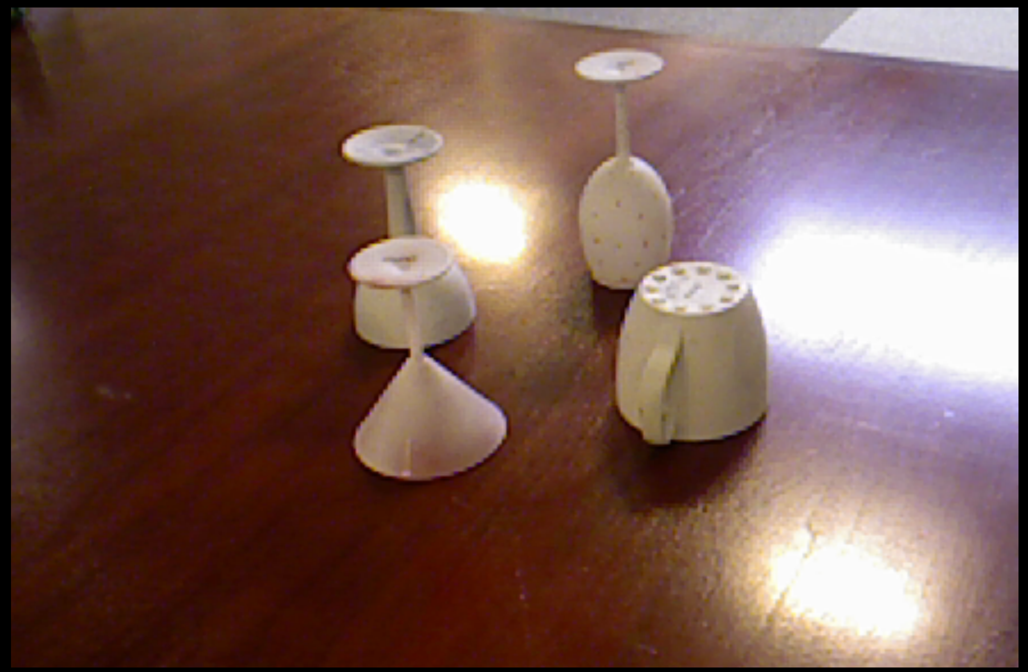
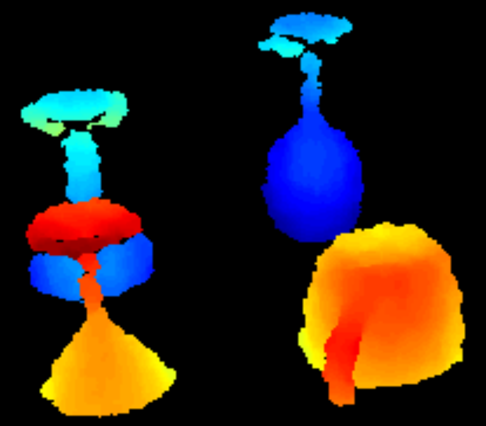
Optimize over space of all possible renderings

combinatorial search space

e.g.: 4 objects, 10 (x,y) locations, 10 orientations

100^4 scenes ~12 days @ 10 ms / render on single GPU

PERCH: Perception via Search

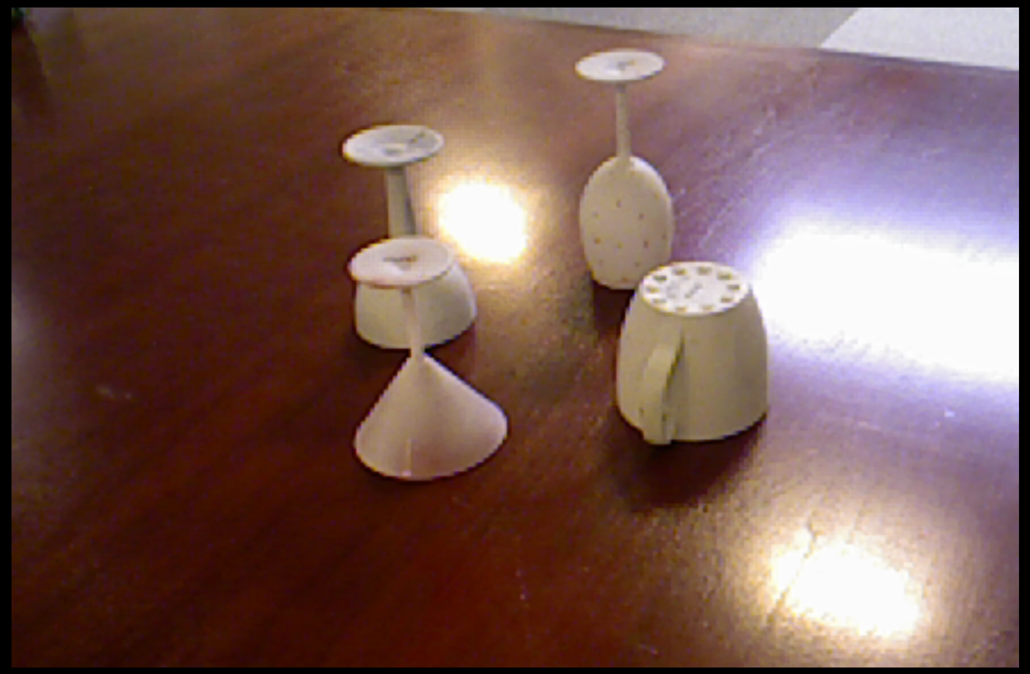
 

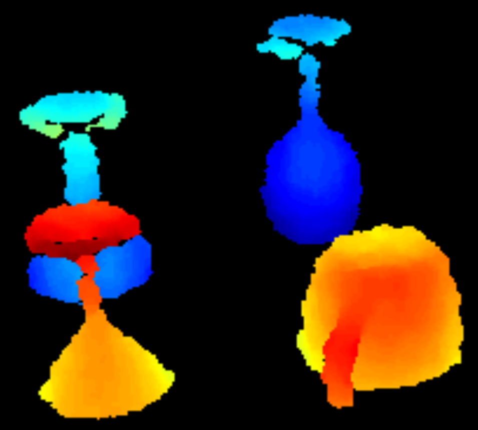
Optimize over space of all possible renderings 

Key Insight

Cast as tree search over
individual object poses 

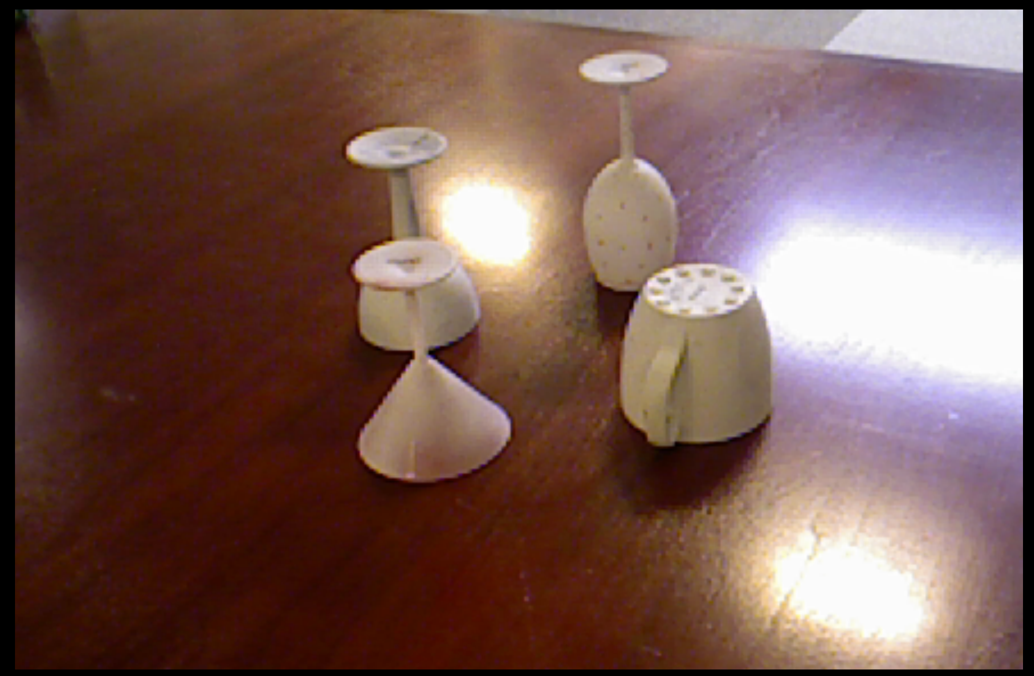
PERCH: Perception via Search

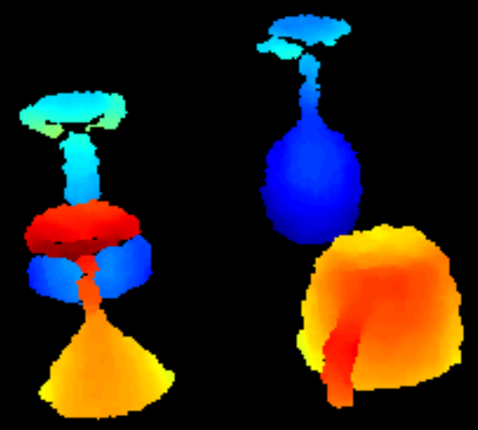




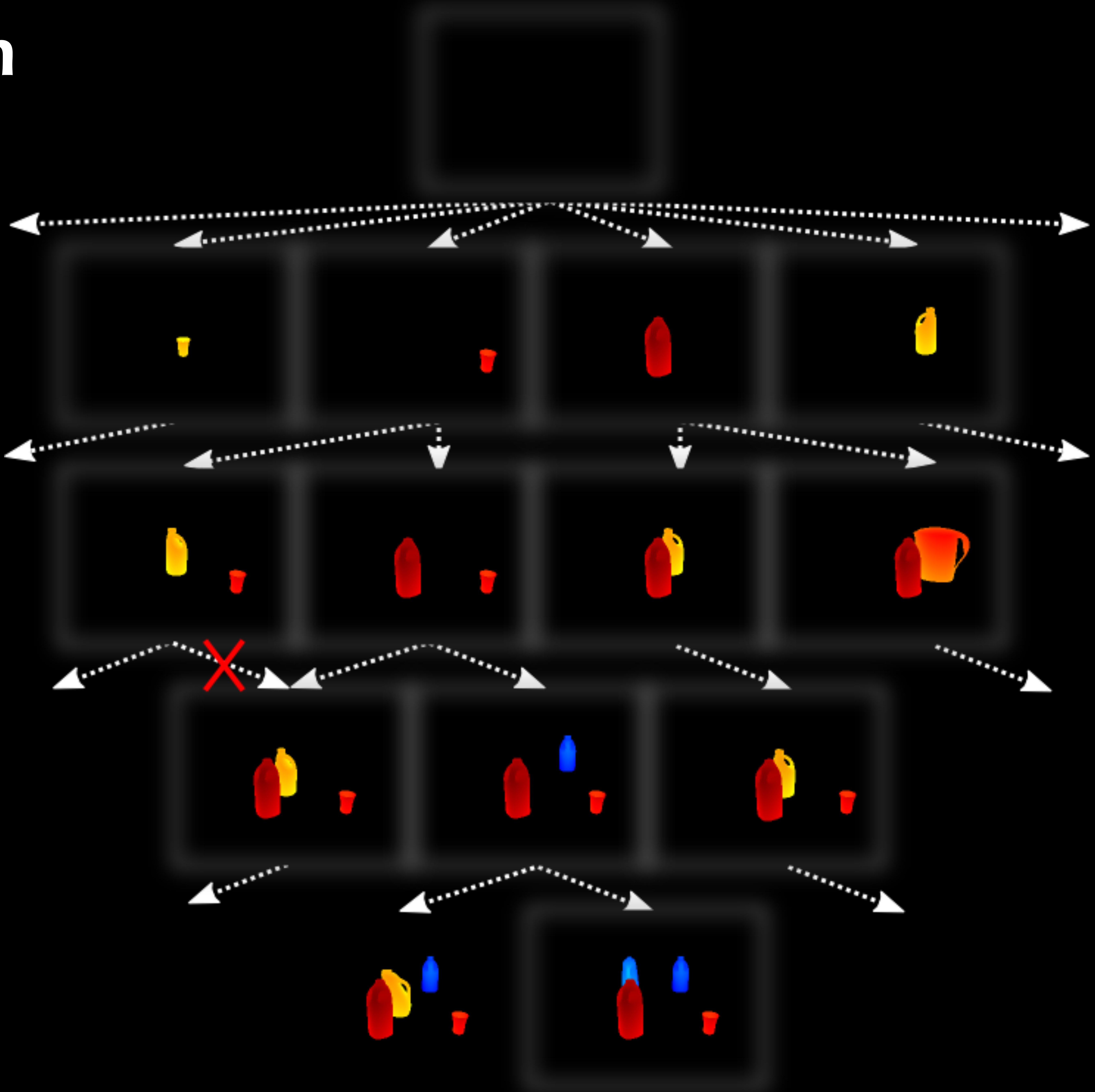
Cast as tree search over
individual object poses

PERCH: Perception via Search

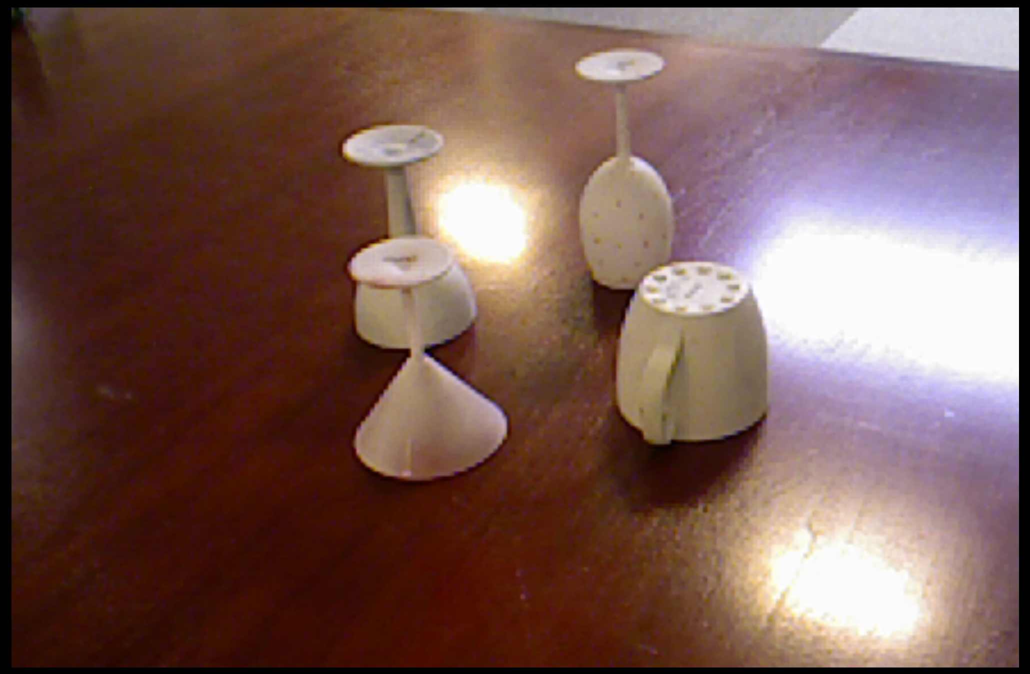


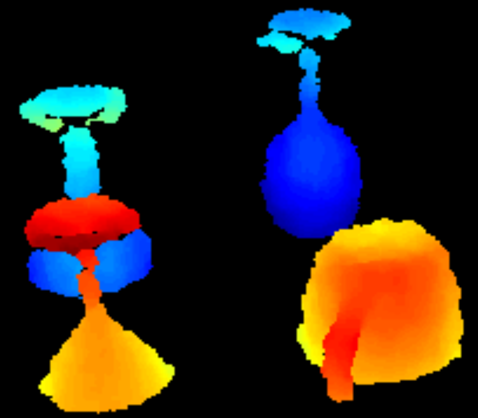


Cast as tree search over
individual object poses

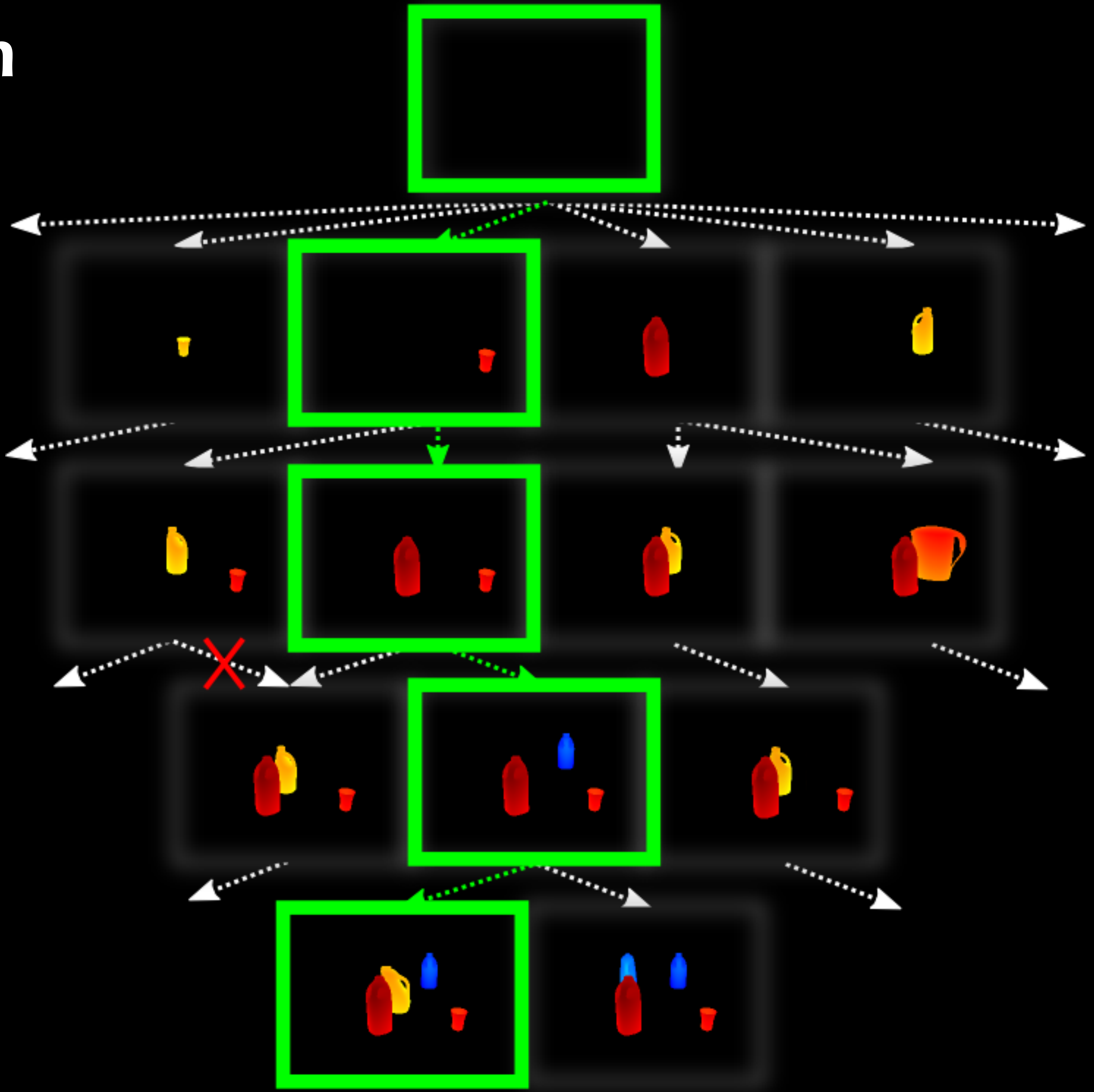


PERCH: Perception via Search



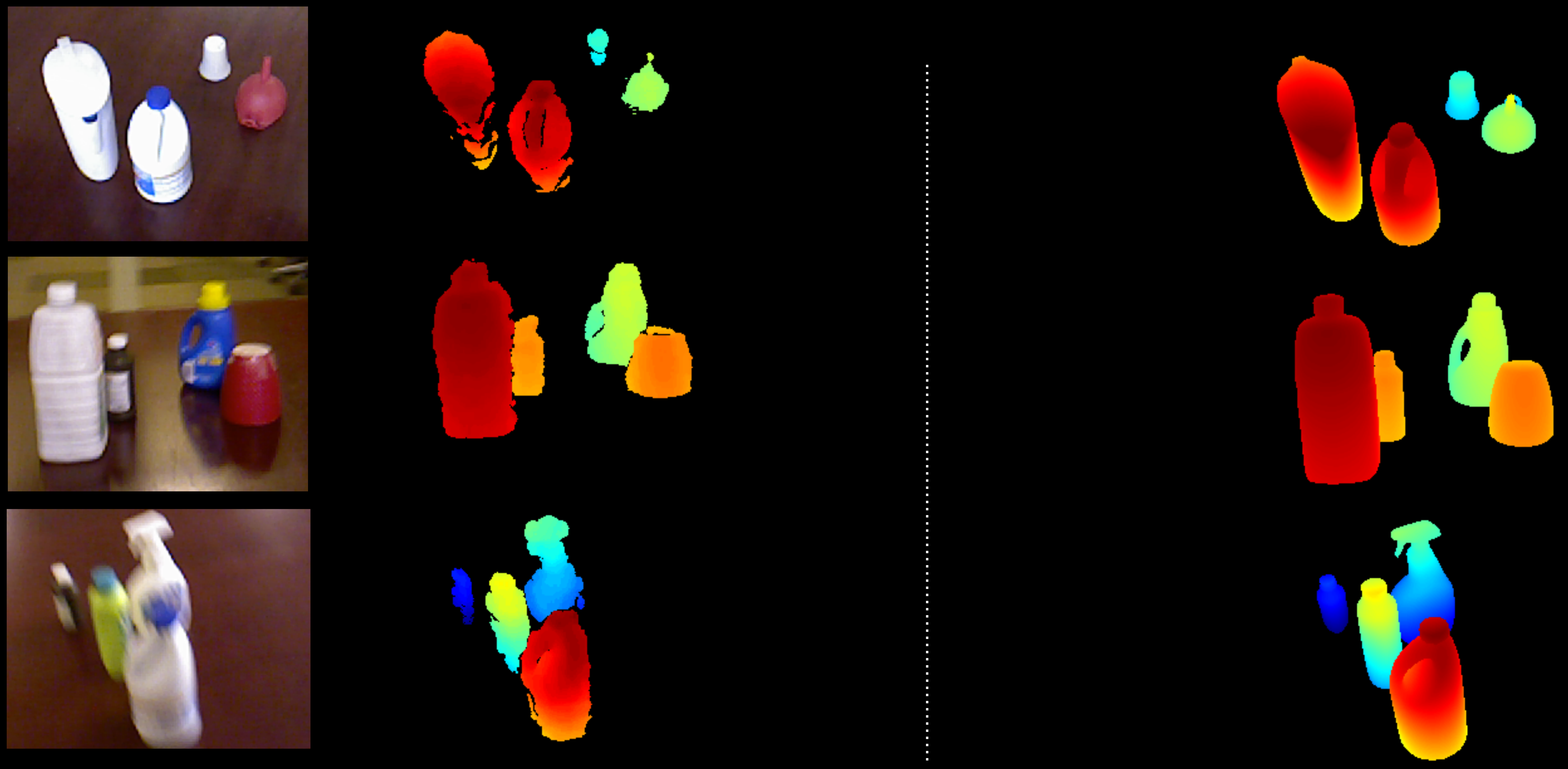


Cast as tree search over *individual* object poses



Experiments: Occlusion Dataset

Input RGB-D | PERCH Output



Experiments: Occlusion Dataset

Input RGB-D

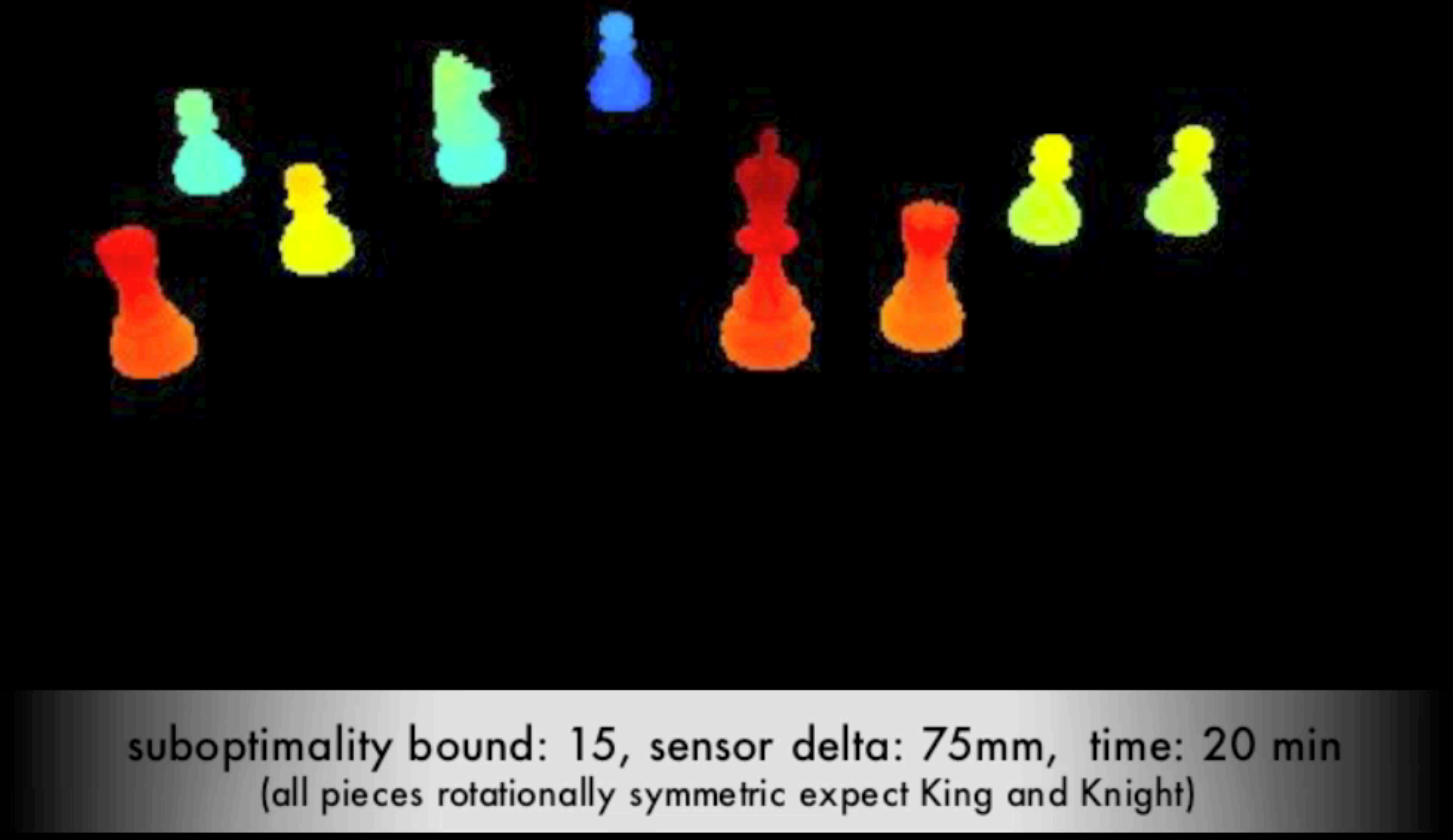
PERCH Output

Outperforms brute-force ICP and OUR-CVFH baselines

Experiments: Scaling

Input Depth Image

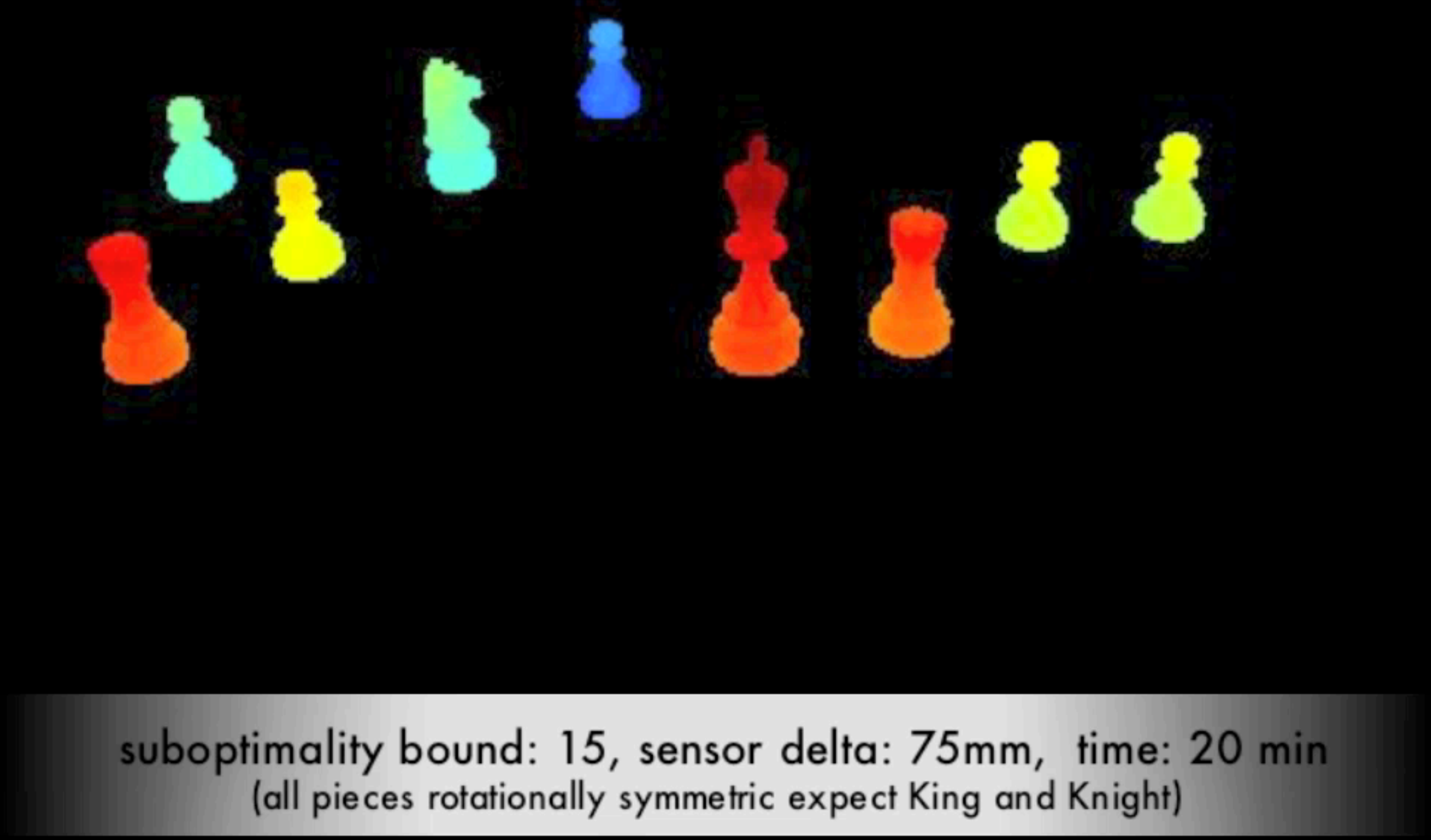
Output Depth Image



Experiments: Scaling

Input Depth Image

Output Depth Image



Summary

- 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

Current Work

Statistical learners as discriminative heuristics for search

Discriminatively-guided Deliberative Perception

Robotics: Science and Systems (RSS) '16

RCNN-Heuristics

code: github.com/venkatrn/perception